

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 138.669 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-F

Perfect score: 7276

Sequence: 1 MAWKTLPIYLLLSVFVIQ.....ARALTTRSGQTLSKVWNCP 1363

Scoring table: BIOSIM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7276	100.0	1363	13	US-10-124-557-52
2	7261.9	99.8	1404	9	Sequence 52, Appl
3	7261.9	99.8	1404	11	Sequence 30, Appl
4	7261.9	99.8	1404	11	US-09-897-188-1
5	7261.9	99.8	1404	13	Sequence 1, Appl
6	7261.9	99.8	1404	13	US-10-124-557-2
7	7058.7	97.0	1320	13	Sequence 2, Appl
8	7058.7	97.0	1320	13	US-10-124-557-62
9	7044.6	96.8	1361	13	Sequence 62, Appl
10	6987.9	96.0	1313	13	Sequence 46, Appl
11	6878.9	94.5	1314	13	Sequence 60, Appl
12	6794.7	93.4	1270	13	Sequence 40, Appl
13	6780.6	93.2	1311	13	Sequence 142, Appl
14	5820.9	80.0	1140	13	Sequence 48, Appl
					Sequence 50, Appl
					Sequence 44, Appl
					Sequence 42, Appl
					Sequence 104, Appl

15	5561	76.4	1049	13	US-10-124-557-58	Sequence 58, Appl
16	5495.5	75.5	1038	13	US-10-124-557-74	Sequence 74, Appl
17	5231.9	71.9	1022	13	US-10-124-557-84	Sequence 84, Appl
18	5011	68.9	941	13	US-10-124-557-14	Sequence 14, Appl
19	3897.7	53.6	792	9	US-09-802-207-27	Sequence 27, Appl
20	2850.9	39.2	538	14	US-10-038-694-3	Sequence 3, Appl
21	2160.9	29.7	422	13	US-10-124-557-68	Sequence 68, Appl
22	2146.8	29.5	463	13	US-10-124-557-54	Sequence 54, Appl
23	2032.8	27.9	423	13	US-10-124-557-66	Sequence 66, Appl
24	1886.9	25.9	372	13	US-10-124-557-64	Sequence 64, Appl
25	1714.7	23.6	401	9	US-09-802-207-29	Sequence 29, Appl
26	1401.6	19.3	292	16	US-10-468-910-4	Sequence 4, Appl
27	1371	18.8	5179	9	US-08-922-217-1068	Sequence 1068, Ap
28	1371	18.8	5179	9	US-09-833-263-1068	Sequence 1068, Ap
29	1371	18.8	5179	13	US-10-025-380-1068	Sequence 1068, Ap
30	1371	18.8	5179	16	US-10-734-564-121	Sequence 121, App
31	1200.1	16.5	296	13	US-10-124-557-70	Sequence 70, Appl
32	992.9	13.6	1460	14	US-10-295-027-428	Sequence 128, App
33	983	13.5	1367	9	US-09-801-368-108	Sequence 108, App
34	968.5	13.3	6642	14	US-10-369-493-5013	Sequence 5013, Ap
35	965.1	13.3	3507	14	US-10-369-493-5784	Sequence 5784, Ap
36	956.9	13.2	1325	9	US-09-864-761-35612	Sequence 35612, A
37	953.7	13.1	19723	15	US-10-084-846A-5	Sequence 5, Appl
38	946.2	13.0	237	13	US-10-124-557-72	Sequence 72, Appl
39	941.9	12.9	5935	14	US-10-243-243A-8	Sequence 8, Appl
40	938.6	12.9	5877	14	US-10-142-515-11	Sequence 11, Appl
41	937.1	12.9	22152	16	US-10-715-066-5	Sequence 5, Appl
42	923.9	12.7	220	13	US-10-124-557-96	Sequence 96, Appl
43	912	12.5	2090	16	US-10-408-765A-2318	Sequence 2318, Ap
44	897.3	12.3	4322	16	US-10-437-963-104793	Sequence 104793, Ap
45	895.7	12.3	2971	14	US-10-146-473-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1

US-10-124-557-52

; Sequence 52, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

APPLICANT: Turner, Katherine C.

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

QY 26 -----ELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 79
Db 61 KRVCCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
QY 80 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 139
Db 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 180
QY 140 KIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199
Db 181 KIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240
QY 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKOTSDG 259
Db 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKOTSDG 300
QY 260 KEKTTSAKETQSIKTSKADLAPTSQVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 319
Db 301 KEKTTSAKETQSIKTSKADLAPTSQVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 360
QY 320 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 379
Db 361 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 420
QY 380 APITTKSAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 439
Db 421 APITTKSAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 480
QY 440 EPAPTAPKAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 499
Db 481 EPAPTAPKAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 540
QY 500 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPT 559
Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPT 600
QY 560 APTAPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPT 619
Db 601 APTAPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPT 660
QY 620 PEPAETTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPT 679
Db 661 PEPAETTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKAPTTTPKEPAPT 720
QY 680 APTTPKAPKAPKELAPTTTPKEPAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAP 739
Db 721 APTTPKAPKAPKELAPTTTPKEPAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAP 780
QY 740 TAPTTLKEPAPTTTPKAPKAPKELAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAP 799
Db 781 TAPTTLKEPAPTTTPKAPKAPKELAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAP 840
QY 800 KPAPTTPETPPPTTSVSTPTTKETTTTHKSPDESTPELSAPPTPKALENSKPEGVPT 859
Db 841 KPAPTTPETPPPTTSVSTPTTKETTTTHKSPDESTPELSAPPTPKALENSKPEGVPT 900
QY 860 TKTPAATKEMTTTAKDKTTERDLRTTPETTTAAAPKWKETAITTEKTTESKITATTQV 919
Db 901 TKTPAATKEMTTTAKDKTTERDLRTTPETTTAAAPKWKETAITTEKTTESKITATTQV 960
QY 920 TSTTTQDTPPKITLTKTTLLAPKVTITTKKLTITTEIMNKPEETAKPKDRATNSKATTPK 979
Db 961 TSTTTQDTPPKITLTKTTLLAPKVTITTKKLTITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 980 POKPTKAPKPTSTKPKTMPVRKDKTTTPFRKMTSTWPELNPTSRIAEAMLQTTTRN 1039
Db 1021 POKPTKAPKPTSTKPKTMPVRKDKTTTPFRKMTSTWPELNPTSRIAEAMLQTTTRN 1080
QY 1040 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVEPEVTPDMDYLPRVPOGIIINPMLS 1099
Db 1081 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVEPEVTPDMDYLPRVPOGIIINPMLS 1140
QY 1100 DETNICNGKPVLDGLTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTFT 1159

Db 1141 DETNICNGKPVLDGLTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTFT 1200
QY 1160 RCNCEGKTFKQYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYKNWPESVY 1219
Db 1201 RCNCEGKTFKQYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1220 FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYEMTQVRRRRFERAIGPSQTHIRIQY 1279
Db 1261 FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYEMTQVRRRRFERAIGPSQTHIRIQY 1320
QY 1280 SPARLAYODKGVHLNWKVSVILNWKGLNVTSAISLNRKPDGYDYAFSKDQYNNIDV 1339
Db 1321 SPARLAYODKGVHLNWKVSVILNWKGLNVTSAISLNRKPDGYDYAFSKDQYNNIDV 1380
QY 1340 PSRTARAITTRSGOTLSKVMWNC 1363
Db 1381 PSRTARAITTRSGOTLSKVMWNC 1404

RESULT 3
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

Query Match 99.8%; Score 7261.9; DB 11; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQ-----25
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATCNCYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 79
Db 61 KRVCCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
QY 80 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 139
Db 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 180
QY 140 KIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199
Db 181 KIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240
QY 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKOTSDG 259
Db 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKOTSDG 300
QY 260 KEKTTSAKETQSIKTSKADLAPTSQVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 319
Db 301 KEKTTSAKETQSIKTSKADLAPTSQVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 360
QY 320 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 379
Db 361 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 420

260 KEKTTSAKETOSIEKTSKVLAKPTPKAETTTKGPALTTKBPPTTKBPAS 319
Db KEKTTSAKETOSIEKTSKVLAKPTPKAETTTKGPALTTKBPPTTKBPAS 360
320 TTPKEPTPTTKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBP 379
Db TTPKEPTPTTKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBP 420
380 APTTTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTPK 439
Db APTTTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTPK 480
440 EPAPTAPKBPATTPKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 499
Db EPAPTAPKBPATTPKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 540
500 TTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 559
Db TTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 600
560 APTAPKBPATTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTP 619
Db APTAPKBPATTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTP 660
620 PBPAPTTPKAAAPNTPKBPATTPKBPATTTKBPATTTKBPATTTKBP 679
Db PBPAPTTPKAAAPNTPKBPATTTKBPATTTKBPATTTKBPATTTKBP 720
680 APTPKKAPKELAPTTTKEPTSTTSKBPATTTKBPATTTKBPATTTKBP 739
Db APTPKKAPKELAPTTTKEPTSTTSKBPATTTKBPATTTKBPATTTKBP 780
740 TAPTTLKEBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTTKBPAT 799
Db TAPTTLKEBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTTKBPAT 840
800 KPAPTTPPTPTTSVSTPTTKBTTHKSPDESTPELSAPTTPKALENSPK 859
Db KPAPTTPPTPTTSVSTPTTKBTTHKSPDESTPELSAPTTPKALENSP 900
860 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAFTTEKTESKITATTQV 919
Db TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAFTTEKTESKITATTQV 960
920 TSTTTQDTPPKLTTLKTTTLAPKVTTKTITTTIMNKPSTAKPKDRATNSKATTPK 979
Db TSTTTQDTPPKLTTLKTTTLAPKVTTKTITTTIMNKPSTAKPKDRATNSKATTPK 1020
980 POKPTKAPKPTSTTKPKTTPVRKPKTTPTRKMTSTMPKNTSRIAEAMLQTTTRN 1039
Db POKPTKAPKPTSTTKPKTTPVRKPKTTPTRKMTSTMPKNTSRIAEAMLQTTTRN 1080
1040 QTPNSKLVBNPKSBDAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1099
Db QTPNSKLVBNPKSBDAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1140
1100 DETNLCNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGISPIDTVPT 1159
Db DETNLCNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGISPIDTVPT 1200
1160 RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPKTFKGFGLTGQIVAAALSTAKYNWPSVY 1219
Db RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPKTFKGFGLTGQIVAAALSTAKYNWPSVY 1260
1220 FFKRGSGIQOYIKQEPVQKCFRRPALNYPVYGMTQVRRRRFRAIGPSQTHIRIQY 1279
Db FFKRGSGIQOYIKQEPVQKCFRRPALNYPVYGMTQVRRRRFRAIGPSQTHIRIQY 1320
1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYVNDV 1339
Db SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYVNDV 1380

1340 PSRTARAITTRSGOTLSKVWYNCP 1363
Db PSRTARAITTRSGOTLSKVWYNCP 1404
RESULT 5
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 99.8%; Score 7261.9; DB 13; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPTLYLLLSVFIQVSSQ----- 25
Db 1 MAWKTLPTLYLLLSVFIQVSSQ----- 25
QY 26 -----ELSCGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 60
Db 61 KRVCCTAELSCGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSPKPNKKTKTKVIESEETIEHVSQENQESSSSSSSSSSSTIW 139
Db 121 PPSGASQTIKSTTKRSPKPNKKTKTKVIESEETIEHVSQENQESSSSSSSSSSSTIW 180

QY 61 SFCAEYHNTPSPSSKAPPPGASQTIKSTTKRSPKPNKXKTKVIESEETEBHSVS 120
DB 61 SFCAEYHNTPSPSSKAPPPGASQTIKSTTKRSPKPNKXKTKVIESEETEBHSVS 115
QY 121 ENQESSSSSSSSSTIWKISKSNSAANRELQKKLVKDNKKNRKXKTKPPVWDE 180
DB 116 -----VKDNKKNRKXKTKPPVWDE 137
QY 181 AGSLONGDFKVTTPOTSTTQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKE 240
DB 138 AGSLONGDFKVTTPOTSTTQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKE 197
QY 241 TVVETKTTTTNNKQSTGDKKETSASKEISAKIAPSKVLAKETPKAETTKG 300
DB 198 TVVETKTTTTNNKQSTGDKKETSASKEISAKIAPSKVLAKETPKAETTKG 257
QY 301 PALTTKEPTTPKEPASSTPKETPTTTSKAPTTPKEPATTTTKSAPTTTKEPATTTT 360
DB 258 PALTTKEPTTPKEPASSTPKETPTTTSKAPTTPKEPATTTTKSAPTTTKEPATTTT 317
QY 361 KEPAATTPKEPATTTTKEPATTTTTSKAPTTPKEPATTTPKKPAPTTPKEPATTTT 420
DB 318 KEPAATTPKEPATTTTKEPATTTTTSKAPTTPKEPATTTPKKPAPTTPKEPATTTT 377
QY 421 TTPKEPATTTKEPATTTPKGPAPTTPKAPAPTTPKEPATTTTKEPATTTTKEPATTTT 480
DB 378 TTPKEPATTTKEPATTTPKGPAPTTPKAPAPTTPKEPATTTTKEPATTTTKEPATTTT 437
QY 481 PAPTTTASAPTTTKEPATTTTTSKAPTTPKPSPTTKEPATTTTKEPATTTTKEPATTTT 540
DB 438 PAPTTTASAPTTTKEPATTTTTSKAPTTPKPSPTTKEPATTTTKEPATTTTKEPATTTT 497
QY 541 KEPAATTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 600
DB 498 KEPAATTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 557
QY 601 PTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 660
DB 558 PTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 617
QY 661 KETAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 720
DB 618 KETAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 677
QY 721 KEPAATTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 780
DB 678 KEPAATTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 737
QY 781 TPKETAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 840
DB 738 TPKETAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 797
QY 841 AEPTPKALENSKPGVPTTTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPMTKET 900
DB 798 AEPTPKALENSKPGVPTTTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPMTKET 857
QY 901 ATTTEKTTESKITATTQVSTTTQDTTTPFKITTLTKTTLAPKVTTKTITTTIMNKP 960
DB 858 ATTTEKTTESKITATTQVSTTTQDTTTPFKITTLTKTTLAPKVTTKTITTTIMNKP 917
QY 961 BETAKPKDRATNSKATTPKPKQKPTKAPKPTSTTKPKTMVPRVKPTTTPPKMTSTTPE 1020
DB 918 BETAKPKDRATNSKATTPKPKQKPTKAPKPTSTTKPKTMVPRVKPTTTPPKMTSTTPE 977
QY 1021 LNPTSRIAEAMLOTTTRNOTPNSKLVNPKSEDAGGAGETPHMLLRBHVMEVETPD 1080
DB 978 LNPTSRIAEAMLOTTTRNOTPNSKLVNPKSEDAGGAGETPHMLLRBHVMEVETPD 1037
QY 1081 MDYLPRVNOGIIINPMLSDETNICNGKPVDLGTLTLRNGTLVAFRGHVFWMLSPPSPSP 1140
DB 1038 MDYLPRVNOGIIINPMLSDETNICNGKPVDLGTLTLRNGTLVAFRGHVFWMLSPPSPSP 1097

QY 1141 ARRITEVWGIBSPIDTFTVTRCNCCEKGTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1200
DB 1098 ARRITEVWGIBSPIDTFTVTRCNCCEKGTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1260
DB 1158 QIVAAALSTAKYKNWPESVYFFFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1217
QY 1261 RRPERAIGSQTHITRIQYSPARLAYQDKGVLNEVKVLSILWGLPNVVTSAISLPNIRK 1320
DB 1218 RRPERAIGSQTHITRIQYSPARLAYQDKGVLNEVKVLSILWGLPNVVTSAISLPNIRK 1277
QY 1321 PDGVDYAFSKDOYVNDVPSRTARALTTRSGOTLSKVWYNCP 1363
DB 1278 PDGVDYAFSKDOYVNDVPSRTARALTTRSGOTLSKVWYNCP 1320

RESULT 7

US-10-124-557-60
; Sequence 60, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; City: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 97.0%; Score 7058.7; DB 13; Length 1320;
Best Local Similarity 96.8%; Pred. No. 2e-162;
Matches 1320; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFBSFERGRCDCAOCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFBSFERGRCDCAOCKKYDKCCPDYE 60
QY 61 SFCAEVHNPSPSSKKAPPPGASGQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
Db 61 SFCAEVHNPSPSSKKAPPPGASGQTIKSTTKRSPKPNKKTKKVIIESEBITE 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKKNRKKKPTPKPPVUDE 180
Db 116 -----VKONKKNRKKKPTPKPPVUDE 137
QY 181 AGSLDNGDFKVTTPDTSTTQHNKVSTSPKIITAKPINRPSLPNSDTSKETSITVNKE 240
Db 138 AGSLDNGDFKVTTPDTSTTQHNKVSTSPKIITAKPINRPSLPNSDTSKETSITVNKE 197
QY 241 TTVETKEITTNKQSTDGKKTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKG 300
Db 198 TTVETKEITTNKQSTDGKKTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKG 257
QY 301 PALATTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSATTPKEPAPTTT 360
Db 258 PALATTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSATTPKEPAPTTT 317
QY 361 KEPAATTPKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTTPKEPAPTTTPKEPTP 420
Db 318 KEPAATTPKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTTPKEPAPTTTPKEPTP 377
QY 421 TTPKEPAPTTKEPAPTTKEPAPAPKPAKPAATTPKEPAPTTPKKPAPTTTPKEPSPTTKE 480
Db 378 TTPKEPAPTTKEPAPTTPKAPAPKPAATTPKEPAPTTTPKEPAPTTTPKEPSPTTKE 437
QY 481 PAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTKEPAPTTTPKEPAPTTPKKPAPTT 540
Db 438 PAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTKEPAPTTTPKEPAPTTPKKPAPTT 497
QY 541 KEPAATTPKEPAPTTTKKPAPAPKPAATTPKEPAPTTPKKLAPTTPEKLAPTTPEKPA 600
Db 498 KEPAATTPKEPAPTTTKKPAPAPKPAATTPKEPAPTTPEKAPTTPEKLAPTTPEKPA 557
QY 601 PTTPEELAPTTPEEPTTTTPEEPAPTTPKAAAPNTPEPAPTTPKKPAPTTTPKEPAPTT 660
Db 558 PTTPEELAPTTPEEPTTTTPEEPAPTTPKAAAPNTPEPAPTTTPKEPAPTTTPKEPAPTT 617
QY 661 KETAPTTPKGAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPT 720
Db 618 KETAPTTPKGAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPT 677
QY 721 PKEPAPTTKEPAPTTPKGAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPT 780
Db 678 PKEPAPTTKEPAPTTPKGAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPT 737
QY 781 TPKEPAPTTKEPAPTTPKAPPTTTPPTTSEVSTPTTKEPTTIHKSPDESTBELS 840
Db 738 TPKEPAPTTKEPAPTTPKAPPTTTPPTTSEVSTPTTKEPTTIHKSPDESTBELS 797
QY 841 AEPTPKALENSKEPVGPTTKTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKET 900
Db 798 AEPTPKALENSKEPVGPTTKTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKET 857
QY 901 ATTTEKTESKIATTTQVSTTQDTPPKITLTKTTLAPKVTITTKITTEIMNKP 960
Db 858 ATTTEKTESKIATTTQVSTTQDTPPKITLTKTTLAPKVTITTKITTEIMNKP 917
QY 961 EETAKPKDRATNSKATTPKQKPTKAPKPTTKPKTMRVKNPKTTPPRKMTSTMPE 1020
Db 918 EETAKPKDRATNSKATTPKQKPTKAPKPTTKPKTMRVKNPKTTPPRKMTSTMPE 977
QY 1021 LNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSDAGAGETPHMLLRPHVFWPEVTPD 1080
Db 978 LNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSDAGAGETPHMLLRPHVFWPEVTPD 1037

QY 1081 MDYLPRVNOGIIINPMLSDETNI CNKGPVDGLTTLRNGTLYAFRGHYFWMLSPPSPSP 1140
Db 1038 MDYLPRVNOGIIINPMLSDETNI CNKGPVDGLTTLRNGTLYAFRGHYFWMLSPPSPSP 1097
QY 1141 ARRI TEVMGIPSPIDTVTRCNCCKTFFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTG 1200
Db 1098 ARRI TEVMGIPSPIDTVTRCNCCKTFFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVQKCPGRRPALNYPVYGMTQVRR 1260
Db 1158 QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVQKCPGRRPALNYPVYGMTQVRR 1217
QY 1261 RRFERAI GSPHTTIRIOYSPARLAYQDKGLHNVKYSIILWRGLPNVVTSAISLPINRK 1320
Db 1218 RRFERAI GSPHTTIRIOYSPARLAYQDKGLHNVKYSIILWRGLPNVVTSAISLPINRK 1277
QY 1321 PDGYDYAFSKDQYNNIDVPSRTARAITTRSGQTL SKVWYNCP 1363
Db 1278 PDGYDYAFSKDQYNNIDVPSRTARAITTRSGQTL SKVWYNCP 1320

RESULT 8
US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match			
Best Local Similarity 96.8%; Score 7044.6; DB 13; Length 1361;			
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;			
Qy	1	MAWKTLPYVLLLLSVFVIQVSSQ-----	25
Db	1	MAWKTLPYVLLLLSVFVIQVSSQDLSSCAGRGEGYSROATCNCYDQHYMECCPDF	60
Qy	26	-----ELSCGRCPESFERGECDCQCKYDKCCDDYSCFAEVNPTSPPSKKAP	79
Db	61	KRVCTAELSCGRCPESFERGECDCQCKYDKCCDDYSCFAEVNPTSPPSKKAP	120
Qy	80	PPSGASQTIKSTTKSPKPPNKKTKVIESEIEETEEHSVSENQESSSSSSSSSIW	139
Db	121	PPSGASQTIKSTTKSPKPPNKKTKVIESEITE-----	156
Qy	140	KIKSKNSAANRELQKULVKONKKNRTKKXPTPKPPVVDVBAAGSLDNGDFKVTTPDTST	199
Db	157	-----VKDNKKNRTKKXPTPKPPVVDVBAAGSLDNGDFKVTTPDTST	197
Qy	200	TOHNKSTSPKLTAKPINRPSLPNSDTSKETSLTVNKEITVETKETTNNKQTSIDG	259
Db	198	TOHNKSTSPKLTAKPINRPSLPNSDTSKETSLTVNKEITVETKETTNNKQTSIDG	257
Qy	260	KEKTTSAKETQSIKBTSAKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS	319
Db	258	KEKTTSAKETQSIKBTSAKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS	317
Qy	320	TTPEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTKEPATTTTKEPATTTKEP	379
Db	318	TTPEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTKEPATTTTKEPATTTKEP	377
Qy	380	APTTTKSAPTTPKEPATTPKKPATTPKEPATTPKEPTTTPKEPATTTKEPATTPK	439
Db	378	APTTTKSAPTTPKEPATTPKKPATTPKEPATTPKEPTTTPKEPATTTKEPATTPK	437
Qy	440	EPAPTAPKKPATTPKEPATTPKEPATTTTKESPPTTPKEPATTTTKSAPTTTKEPAT	499
Db	438	EPAPTAPKKPATTPKEPATTTTKESPPTTPKEPATTTTKSAPTTTKEPATTTKEPAT	497
Qy	500	TTKSAPTTPKEPSPTTTPKEPATTPKGPATTPKGPATTPKGPATTTTPKEPATTTTKP	559
Db	498	TTKSAPTTPKEPSPTTTPKEPATTPKGPATTPKGPATTPKGPATTTTPKEPATTTKP	557
Qy	560	APTAPKEPATTPKETAPTTPKLTPTTPEKLAPTTPEKLAPTTPEKLAPTTPEEPTPTT	619
Db	558	APTAPKEPATTPKETAPTTPKLTPTTPEKLAPTTPEKLAPTTPEKLAPTTPEEPTPTT	617
Qy	620	PEEPAPTTPKAAAPTTPKEPATTPKEPATTPKEPATTPKGTAPTTPKGTAPTTLKEP	679
Db	618	PEEPAPTTPKAAAPTTPKEPATTPKEPATTPKEPATTPKGTAPTTPKGTAPTTLKEP	677
Qy	680	APTTPKPAKELAPTTPKEPTSTSDKPAPTTPKGTAPTTPKGPATTPKEPATTPKGP	739
Db	678	APTTPKPAKELAPTTPKEPTSTSDKPAPTTPKGTAPTTPKGPATTPKEPATTPKGP	737
Qy	740	TAPTTLKEPATTPKPAKELAPTTPKGTSTSDKPAPTTPKGTAPTTPKGTAPTTPKEPATTPK	799
Db	738	TAPTTLKEPATTPKPAKELAPTTPKGTSTSDKPAPTTPKGTAPTTPKGTAPTTPKEPATTPK	797
Qy	800	KPATTPPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKEPGVPT	859
Db	798	KPATTPPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKEPGVPT	857
Qy	860	TKTPAATKPEMTTAKDKTTERDLRTPTTETTAAPKMKETATTTKTESKITATTQV	919
Db	858	TKTPAATKPEMTTAKDKTTERDLRTPTTETTAAPKMKETATTTKTESKITATTQV	917
Qy	920	TSTTTQDTTTPFKLTTLKTTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK	979
Db	918	TSTTTQDTTTPFKLTTLKTTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK	977

Qy	980	POKPTKADKKSTSTKPKTMRVRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLOTTTRN	1039
Db	978	POKPTKADKKSTSTKPKTMRVRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLOTTTRN	1037
Qy	1040	QTPNSKLVENPKSEBDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS	1099
Db	1038	QTPNSKLVENPKSEBDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS	1097
Qy	1100	DETINCNKPGVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT	1159
Db	1098	DETINCNKPGVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT	1157
Qy	1160	RCNCEGKTFRKDSQYWRFTNDIKDAGYPKIFKFGGLTGQIVAAALSTAKYKNWPESVY	1219
Db	1158	RCNCEGKTFRKDSQYWRFTNDIKDAGYPKIFKFGGLTGQIVAAALSTAKYKNWPESVY	1217
Qy	1220	FFKGGSTQQYIYKQEPVQKCPGRRPALNYPVYEMTOVRRRRPERAIGPSQTHIRIQY	1279
Db	1218	FFKGGSTQQYIYKQEPVQKCPGRRPALNYPVYEMTOVRRRRPERAIGPSQTHIRIQY	1277
Qy	1280	SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLNPIRKPDGYDYVAFSKDQYINIDV	1339
Db	1278	SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLNPIRKPDGYDYVAFSKDQYINIDV	1337
Qy	1340	PSRTARAITTSSGOTLSKVMWVNCV 1363	
Db	1338	PSRTARAITTSSGOTLSKVMWVNCV 1361	
RESULT 9			
US-10-124-557-142			
; Sequence 142, Application US/10124557			
; Publication No. US20020137894A1			
; GENERAL INFORMATION:			
; APPLICANT: Turner, Katherine			
; Clark, Stephen C.			
; Jacobs, Kenneth			
; Hewick, Rodney M.			
; Geshner, Thomas G.			
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors			
; NUMBER OF SEQUENCES: 143			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Genetics Institute, Inc.			
; STREET: 87 Cambridgepark Drive			
; CITY: Cambridge			
; STATE: Massachusetts			
; COUNTRY: U.S.A.			
; ZIP: 02140			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patentin Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/10/124,557			
; FILING DATE: 16-Apr-2002			
; CLASSIFICATION: <Unknown>			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/643,502			
; FILING DATE: 18-JAN-1991			
; APPLICATION NUMBER: US 07/546,114			
; FILING DATE: 29-JUN-1990			
; APPLICATION NUMBER: US 07/457,196			
; FILING DATE: 29-DEC-1989			
; APPLICATION NUMBER: US 07/390,901			
; FILING DATE: 08-AUG-1989			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Cseri, Luann			
; REGISTRATION NUMBER: 31,822			
; REFERENCE/DOCKET NUMBER: GI 5190			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (617)876-1170			
; TELEFAX: (617)876-5851			

```
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match          96.2%; Score 7002; DB 13; Length 1313;
Best Local Similarity 96.3%; Pred. No. 4.5e-161;
Matches 1313; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQOVSSQELSCKRCRCSFESRGRECDCAQCKYDKCCPDYE 60
DB 1 MAMKTLPIYLLLSVFIQOVSSQELSCKRCRCSFESRGRECDCAQCKYDKCCPDYE 60
QY 61 SFCAEVHNPTSPSPSKKXAPPSSGASQTIKSTTKRSPKPNKKTKKVIESBEITEHSVS 120
DB 61 SFCA-----EHSVS 70
QY 121 ENQSSSSSSSSSSSIWIKSSKNSAANRELQKLVKDKNKKNRKTKKPTPKPPVVDE 180
DB 71 ENQSSSSSSSSSSSIWIKSSKNSAANRELQKLVKDKNKKNRKTKKPTPKPPVVDE 130
QY 181 AGSGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSLSPNSDTSKETSITVNIKE 240
DB 131 AGSGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSLSPNSDTSKETSITVNIKE 190
QY 241 TVTETKETITTNKOTSDGKEKTTSAKETQIEKTSAMDAPTSKVLAKPTPKAETTTKG 300
DB 191 TVTETKETITTNKOTSDGKEKTTSAKETQIEKTSAMDAPTSKVLAKPTPKAETTTKG 250
QY 301 PALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKGPAPT 360
DB 251 PALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKGPAPT 310
QY 361 KEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAPTTPKGPAPTTTPKEPAPTTKEPTP 420
DB 311 KEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAPTTPKGPAPTTTPKEPAPTTKEPTP 370
QY 421 TTPKEPAPTTKEPAPTTKEPAPTPAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPTTPE 480
DB 371 TTPKEPAPTTKEPAPTTKEPAPTPAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPTTPE 430
QY 481 PAPTTPKSAPTTKEPAPTTKSAPTTPKPSPTTPKEPAPTTTPKEPAPTTTPKGPAPTT 540
DB 431 PAPTTPKSAPTTKEPAPTTKSAPTTPKPSPTTPKEPAPTTTPKEPAPTTTPKGPAPTT 490
QY 541 KEPAPTTKEPAPTTKGPAPTPAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 600
DB 491 KEPAPTTKEPAPTTKGPAPTPAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 550
QY 601 PTTPELAPTTPEEPTPTTPKEPAPTTPKAAAPNTPKGPAPTTTPKEPAPTTTPKEPAPTT 660
DB 551 PTTPELAPTTPEEPTPTTPKEPAPTTPKAAAPNTPKGPAPTTTPKEPAPTTTPKEPAPTT 610
QY 661 KETAPTTKGTAPTTLKGPAPTTPKKAPKELAPTTTPKEPAPTTSDKAPTTPKGTAPT 720
DB 611 KETAPTTKGTAPTTLKGPAPTTPKKAPKELAPTTTPKEPAPTTSDKAPTTPKGTAPT 670
QY 721 PKEPAPTTPKGPAPTTKGTAPTTLKGPAPTTPKKAPKELAPTTTPKGTSTTSKGPAPT 780
DB 671 PKEPAPTTPKGPAPTTKGTAPTTLKGPAPTTPKKAPKELAPTTTPKGTSTTSKGPAPT 730
QY 781 TPKETAPTTPKGPAPTTPKKAPTTTPETPPPTTSVSTPTTPKEPTTIHKSDPSTPBL 840
DB 731 TPKETAPTTPKGPAPTTPKKAPTTTPETPPPTTSVSTPTTPKEPTTIHKSDPSTPBL 790
QY 841 AEPTPKALENSPKSGVPTTKPAATKPEMTTAKDKTERDLRTPETTTAAPTAKMTKET 900
DB 791 AEPTPKALENSPKSGVPTTKPAATKPEMTTAKDKTERDLRTPETTTAAPTAKMTKET 850

RESULT 10
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csert, Luann
; REGISTRATION NUMBER: 31,822
```

```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match          96.0%; Score 6987.9; DB 13; Length 1354;
Best Local Similarity 93.5%; Pred No. 1e-160;
Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLSVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRGEGYRDATCNCYNCQHYMECCPDF 60
QY 26 -----ELSCRGRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db 61 KRVTAEIUSCKRGCFESPERGECDCDAQCKYDKCCPDYESFCA----- 105
QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIEEITEHSHSVSENOESSSSSSSSSSSTI 139
Db 106 -----EHSVSENOESSSSSSSSSSSSSTI 130
QY 140 KIKSKNSAANRELQKLVKNDKNKTKKPTPKPPVVD BAGSLGNDGPKVITPDST 199
Db 131 KIKSKNSAANRELQKLVKNDKNKTKKPTPKPPVVD BAGSLGNDGPKVITPDST 190
QY 200 TQHNKVTSPKLTAKENRPPSLPPNSDTSKETSITVKNKETTVEKETTNNKQSTGD 259
Db 191 TQHNKVTSPKLTAKENRPPSLPPNSDTSKETSITVKNKETTVEKETTNNKQSTGD 250
QY 260 KEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 319
Db 251 KEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 310
QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPE 379
Db 311 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPE 370
QY 380 APTTTKSAPTTPKEPAPTTTKKAPTTTPKEPAPTTTKPEPTTPKEPAPTTTKPE 439
Db 371 APTTTKSAPTTPKEPAPTTTKKAPTTTPKEPAPTTTPKEPTTPKEPAPTTTKPE 430
QY 440 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTKPEPTTKPEPAPTTTKSAPTTKBP 499
Db 431 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTKPEPTTKPEPAPTTTKSAPTTKBP 490
QY 500 TTKSAPTTPKEPPTTKPEPAPTTTPKEPAPTTTKKAPTTTPKEPAPTTTKKP 559
Db 491 TTKSAPTTPKEPPTTKPEPAPTTTPKEPAPTTTPKCAPTTTPKEPAPTTTKKP 550
QY 560 APTAPKEPAPTTTPKETAPTTPKLTTPTTPEKLAPTTTPKEPAPTTPELAPTTPEPTPT 619
Db 551 APTAPKEPAPTTTPKETAPTTPKLTTPTTPEKLAPTTTPKEPAPTTPELAPTTPEPTPT 610
QY 620 PEEPAPTTPKAAAPNTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTKPE 679
Db 611 PEEPAPTTPKAAAPNTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTKPE 670
QY 680 APTTPKPKAPKELAPTTTKPEPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKGP 739
Db 671 APTTPKPKAPKELAPTTTKPEPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKGP 730
QY 740 TAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPTTTPKEPAPTTTPKEPAPTTPK 799
Db 731 TAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPTTTPKEPAPTTTPKEPAPTTPK 790
```

```

QY 800 KPAPTTPTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPGVT 859
Db 791 KPAPTTPTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPGVT 850
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTETTTAKPKMTKETATTTTETTESKITATTTQV 919
Db 851 TKTPAATKPEMTTAKDKTTERDLRTTPTTETTTAKPKMTKETATTTTETTESKITATTTQV 910
QY 920 TSTTTQDTPPKIITLTKTTILAPKVTITTKITITTEIMNKPEETAKEPKDRATNSKATTPK 979
Db 911 TSTTTQDTPPKIITLTKTTILAPKVTITTKITITTEIMNKPEETAKEPKDRATNSKATTPK 970
QY 980 POKTKAPKKTSTKPKTMPRVKPKKPTTPRKMVTSTMPELNPTSRIAEAMLOTTTRPN 1039
Db 971 POKTKAPKKTSTKPKTMPRVKPKKPTTPRKMVTSTMPELNPTSRIAEAMLOTTTRPN 1030
QY 1040 QTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099
Db 1031 QTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1090
QY 1100 DETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1159
Db 1091 DETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1150
QY 1160 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKNWPSVY 1219
Db 1151 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKNWPSVY 1210
QY 1220 FFKRGSGSIQYIYKQEPVQKCPGRRPALNVPVYGMTQVRRRERAIKPSQTHIRIQY 1279
Db 1211 FFKRGSGSIQYIYKQEPVQKCPGRRPALNVPVYGMTQVRRRERAIKPSQTHIRIQY 1270
QY 1280 SPARLAYODKGLVHNEVKVLSILMRGLENVVTSAISLPIRKPDGYDYVAFKDOYXNDV 1339
Db 1271 SPARLAYODKGLVHNEVKVLSILMRGLENVVTSAISLPIRKPDGYDYVAFKDOYXNDV 1330
QY 1340 PSRTARAITTRSGOTLSKQWYNCP 1363
Db 1331 PSRTARAITTRSGOTLSKQWYNCP 1354
```

```

RESULT 11
US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;             Clark, Stephen C.
;             Jacobs, Kenneth
;             Hewick, Rodney M.
;             Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
```

FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50
Query Match 94.5%; Score 6878.9; DB 13; Length 1314;
Best Local Similarity 94.7%; Pred. No. 4.3e-158;
Matches 1292; Conservative 7; Mismatches 14; Indels 51; Gaps 2;
QY 1 MAWKTLPIYLLLSVFVIQOVSSQEL-SCKGRCFESFERGECDCDAQCKKYDKCCPDY 59
DB 1 MAWKTLPIYLLLSVFVIQOVSSQDLSCACGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 60 ESFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPNNKTKKVIIESEETEBHSV 119
DB 61 KRVC-----TAHSV 70
QY 120 SENGESSSSSSSSSTIWKSSKNSAANRELQKLKVNDKNKNTKKPTPKPPVVD 179
DB 71 SENGESSSSSSSSSTIWKSSKNSAANRELQKLKVNDKNKNTKKPTPKPPVVD 130
QY 180 EAGGLONGDKVTPDTSITQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLVNK 239
DB 131 EAGGLONGDKVTPDTSITQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLVNK 190
QY 240 ETVVETKETTNTKQTSIDGKEKTSKAKETOSIEKTSKADLAPTSKVLAKPTPKAETTK 299
DB 191 ETVVETKETTNTKQTSIDGKEKTSKAKETOSIEKTSKADLAPTSKVLAKPTPKAETTK 250
QY 300 GPALTTPKEPTTPPKESPASTPKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTT 359
DB 251 GPALTTPKEPTTPPKESPASTPKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTT 310
QY 360 TKPEAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPT 419
DB 311 TKPEAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPT 370
QY 420 PTTKPEAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTTKPEAPTTTK 479
DB 371 PTTKPEAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTTKPEAPTTTK 430
QY 480 EPAPTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKAPTT 539
DB 431 EPAPTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKAPTT 490
QY 540 PKPEAPTTKPEAPTTTKKAPTTKPEAPTTKPEAPTTKPEAPTTTKKAPTTTKPEK 599
DB 491 PKPEAPTTKPEAPTTTKKAPTTKPEAPTTKPEAPTTKPEAPTTTKKAPTTTKPEK 550
QY 600 APTTPEELAPTTPEPTPTTPEAPTTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 659
DB 551 APTTPEELAPTTPEPTPTTPEAPTTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 610
QY 660 PKETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPT 719
DB 611 PKETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPT 670

QY 720 TPKEAPTTKPEAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAP 779
DB 671 TPKEAPTTKPEAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAP 730
QY 780 TTPKETAPTTKPEAPTTPKKAPAPTTPEPPPTTSEVSTPTTKEPTTIHKSPDESPEL 839
DB 731 TTPKETAPTTKPEAPTTPKKAPAPTTPEPPPTTSEVSTPTTKEPTTIHKSPDESPEL 790
QY 840 SAEPTPKALENSKPEGVPTTKTAPATKPEMTTAKDKTTERDLRTTPTTAAPKMTKE 899
DB 791 SAEPTPKALENSKPEGVPTTKTAPATKPEMTTAKDKTTERDLRTTPTTAAPKMTKE 850
QY 900 TATTTEKTESKITATTQVTSITTTQDTPFKITTTKTTTLLAPKVTITTKITTTTEIMNK 959
DB 851 TATTTEKTESKITATTQVTSITTTQDTPFKITTTKTTTLLAPKVTITTKITTTTEIMNK 910
QY 960 PEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTMPRVRKPKTTPPRKMTSTMP 1019
DB 911 PEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTMPRVRKPKTTPPRKMTSTMP 970
QY 1020 ELNPTSRIAEAMLOTTTRPNOTNSKLVEVNPKSEDAAGAEGETPHMLLRPHVEMPEVTP 1079
DB 971 ELNPTSRIAEAMLOTTTRPNOTNSKLVEVNPKSEDAAGAEGETPHMLLRPHVEMPEVTP 1030
QY 1080 DMDYLPVPVQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPPS 1139
DB 1031 DMDYLPVPVQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPPS 1090
QY 1140 PARRITVWGIPIPIDVTRCNCCEKTFPKDSQYWRFTNDIKDAGYKPKIFKGFGLT 1199
DB 1091 PARRITVWGIPIPIDVTRCNCCEKTFPKDSQYWRFTNDIKDAGYKPKIFKGFGLT 1150
QY 1200 GOIVAAALSTAKYKNWPESVVFYKRGSGSIQOYIYKQEPVQKCPGRRPALNYPVYGMQVR 1259
DB 1151 GOIVAAALSTAKYKNWPESVVFYKRGSGSIQOYIYKQEPVQKCPGRRPALNYPVYGMQVR 1210
QY 1260 RRRFERAIGPSQHTIRIQSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIR 1319
DB 1211 RRRFERAIGPSQHTIRIQSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIR 1270
QY 1320 KPDGYDYAFSKDQYINIDVPSRTARALTTRSGOTLSKVVNCP 1363
DB 1271 KPDGYDYAFSKDQYINIDVPSRTARALTTRSGOTLSKVVNCP 1314
RESULT 12
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Geener, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Juann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

```

```

Query Match      93.4%; Score 6794.7; DB 13; Length 1270;
Best Local Similarity 93.2%; Pred. No. 4.4e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1  MANKTLPIYLLLSVFVIQVSSQELSCKGRCPESFERGRECDCAQCKYDKCCPDYE 60
Db      |||
QY 1  MANKTLPIYLLLSVFVIQVSSQELSCKGRCPESFERGRECDCAQCKYDKCCPDYE 60
Db      |||

QY 61  SFCAEVHNPTSPSSKXKAPPSSGASQITKSTKSPKPPNKKTKVIESEITEHSVS 120
Db      |||
QY 61  SFCAE-----
Db      |||

QY 121  ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLKVKDKNKKRTKKKPTPKPPVVDE 180
Db      -----VKDNKKNRKTKKPTPKPPVVDE 87

QY 181  AGSGLNDGPKVITPTDSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 240
Db      |||
QY 88  AGSGLNDGPKVITPTDSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 147
Db      |||

QY 241  TTVEKETTITNKQTSIDGKEKTSKAKTOSIEKTSKAKLAPTSKVLAKETPKAETTKG 300
Db      |||
QY 148  TTVEKETTITNKQTSIDGKEKTSKAKTOSIEKTSKAKLAPTSKVLAKETPKAETTKG 207
Db      |||

QY 301  PALTTKPEPTTKPEPASTTKPEPTTITKSAPTTKPEPAPTTTKSAPTTKPEPAPTTT 360
Db      |||
QY 208  PALTTKPEPTTKPEPASTTKPEPTTITKSAPTTKPEPAPTTTKSAPTTKPEPAPTTT 267
Db      |||

QY 361  KEPAPTTKPEPAPTTTKPEPAPTTKGAPTTKPEPAPTTPKKAPATTTPKPEPTP 420
Db      |||
QY 268  KEPAPTTKPEPAPTTTKPEPAPTTKGAPTTKPEPAPTTPKKAPATTTPKPEPTP 327
Db      |||

QY 421  TTPKPEPAPTTKPEPAPTTKPEPAPTKKAPATTTPKPEPAPTTTKPEPAPTTKPE 480
Db      |||
QY 328  TTPKPEPAPTTKPEPAPTTKPEPAPTKKAPATTTPKPEPAPTTTKPEPAPTTKPE 387
Db      |||

QY 481  PAPTTKSAPTTTKPEPAPTTTKSAPTTTKPEPSTTTTKPEPAPTTTKPEPAPTTPKK 540
Db      |||
QY 388  PAPTTKSAPTTTKPEPAPTTTKSAPTTTKPEPSTTTTKPEPAPTTTKPEPAPTTPKK 447
Db      |||

QY 541  KEPAPTTKPEPAPTTTKKAPAPKAPKPEPAPTTPKETAPTTPKLTPTTPEKLAFTPEK 600
Db      |||
QY 448  KEPAPTTKPEPAPTTTKKAPAPKAPKPEPAPTTPKETAPTTPKLTPTTPEKLAFTPEK 507
Db      |||

QY 601  PTTPELAPTTPEEPPTTTPPEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTT 660
Db      |||
QY 508  PTTPELAPTTPEEPPTTTPPEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTT 567
Db      |||

```

```

QY 661  KETAPTTKGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTT 720
Db      |||
QY 568  KETAPTTKGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTT 627
Db      |||
QY 721  PKEPAPTTPKBPAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATT 780
Db      |||
QY 628  PKEPAPTTPKBPAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATT 687
Db      |||
QY 781  TPKETAPTTPKBPAPTTPKKAPATTTPPTTSEVSTPTTKEPTTTHKSDESTPELS 840
Db      |||
QY 688  TPKETAPTTPKBPAPTTPKKAPATTTPPTTSEVSTPTTKEPTTTHKSDESTPELS 747
Db      |||
QY 841  AEPTPKALENSPKBPVPTTKTPAATKPEMTTAKDKTTERDLRTPPTTTAAAPKWTKET 900
Db      |||
QY 748  AEPTPKALENSPKBPVPTTKTPAATKPEMTTAKDKTTERDLRTPPTTTAAAPKWTKET 807
Db      |||
QY 901  ATTTEKTTESKITATTTQVTSITTTODTTPPKITLTKTTTLAPKVTTTKTTTTEIMNKP 960
Db      |||
QY 808  ATTTEKTTESKITATTTQVTSITTTODTTPPKITLTKTTTLAPKVTTTKTTTTEIMNKP 867
Db      |||
QY 961  EETAKPKDRATNSKATTPKQKPKAKPKPTSTKKPKTMPRVKPKTTPTPRKMTSTMP 1020
Db      |||
QY 868  EETAKPKDRATNSKATTPKQKPKAKPKPTSTKKPKTMPRVKPKTTPTPRKMTSTMP 927
Db      |||
QY 1021  LNPTSRITAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPD 1080
Db      |||
QY 928  LNPTSRITAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPD 987
Db      |||
QY 1081  MDYLPRVNPNGIILNPMLSDETNI CNKGPVDGLTLLRNGTLVAFRGHYFWMLSFSPSP 1140
Db      |||
QY 988  MDYLPRVNPNGIILNPMLSDETNI CNKGPVDGLTLLRNGTLVAFRGHYFWMLSFSPSP 1047
Db      |||
QY 1141  ARRITEVWGIPSIDTVFTRCNCEGKTFPFKDSQYWRFTINDIKDAGYKPKIFKGFGLTG 1200
Db      |||
QY 1048  ARRITEVWGIPSIDTVFTRCNCEGKTFPFKDSQYWRFTINDIKDAGYKPKIFKGFGLTG 1107
Db      |||
QY 1201  QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1260
Db      |||
QY 1108  QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1167
Db      |||
QY 1261  RRFERAIGPSQTHIRIQSPARLAYQDKGVLHNEKVSVILWGLPNVVTSAISLPNIRK 1320
Db      |||
QY 1168  RRFERAIGPSQTHIRIQSPARLAYQDKGVLHNEKVSVILWGLPNVVTSAISLPNIRK 1227
Db      |||
QY 1321  PDGYDYAFSKQOYINIDVPSRTARAITRSQGTLSKVWYNCP 1363
Db      |||
QY 1228  PDGYDYAFSKQOYINIDVPSRTARAITRSQGTLSKVWYNCP 1270
Db      |||

```

```

RESULT 13
US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;             Clark, Stephen C.
;             Jacobs, Kenneth
;             Hewick, Rodney M.
;             Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42
Query Match 93.2%; Score 6780.6; DB 13; Length 1311;
Best Local Similarity 90.5%; Pred. No. 1e-155;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
1 MAWKTLPIYLLLSVFIQVSSQ----- 25
1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCVDYNCQHMECCPDF 60
26 -----ELCKGRFCFESPERGECDCDAQCKYDKCCPDYEFCAEVNPTSPSSKAP 79
61 KRVTAEELCKGRFCFESPERGECDCDAQCKYDKCCPDYEFCAE----- 106
80 PPGASQTIKTTKRSPPPNKKTKVIESEEITEHHSVENQESSSSSSSSSSSTIW 139
107 ----- 106
140 KIKSSKNSAANRELQKLUKDKNKKNTKKXPTPKPPVVDVAGSLDNGDFKVTTPDST 199
107 -----VKDNKKNTKKXPTPKPPVVDVAGSLDNGDFKVTTPDST 147
200 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKQTSDDG 259
148 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKQTSDDG 207
260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 319
208 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTKPEPAS 267
320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
268 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
380 APTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 439
328 APTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 387
440 BPAPTAKKPAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 499
388 BPAPTAKKPAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 447
500 TTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKPEAPTTTKPE 559
448 TTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKPEAPTTTKPE 507

RESULT 14
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge

```

; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match      80.0%; Score 5820.9; DB 13; Length 1140;
Best Local Similarity 96.4%; Pred. No. 1.3e-132;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIOVSSQ-----25
Db 1 MAWKTLPIYLLLSLVFVIOVSSQDLSSCAGRCGEYSRDATCNDYNCQHMECCPDF 60
QY 26 -----ELSCGRGCFEFGRECDCAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 79
Db 61 KRVCIAELSCGRGCFEFGRECDCAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENGQSSSSSSSSSSSTI 139
Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENGQSSSSSSSSSSSTI 180
QY 140 KIKSSKNSAANRELQKLVKDKNKKRKKKPTKPPVDEAGSLDNGDFKVTTPDTST 199
Db 181 KIKSSKNSAANRELQKLVKDKNKKRKKKPTKPPVDEAGSLDNGDFKVTTPDTST 240
QY 200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETTVEKETTINKQISTDG 259
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETTVEKETTINKQISTDG 300
QY 260 KEKTTSAKETQSIKTSKADLAFTSKVLAKPTPKAETTTKGPALTTPKBPTTTPKEPAS 319
Db 301 KEKTTSAKETQSIKTSKADLAFTSKVLAKPTPKAETTTKGPALTTPKBPTTTPKEPAS 360
QY 320 TTPKEPTPTTIKSAPTTTPKEPARTTKSAPTTTPKEPARTTKBAPTTPKEPARTTKBP 379
Db 361 TTPKEPTPTTIKSAPTTTPKEPARTTKSAPTTTPKEPARTTKBAPTTPKEPARTTKBP 420
QY 380 APTTTKSAPTTTPKEPARTTKBAPTTPKEPARTTKBPTTPKEPARTTKBAPTTPK 439
Db 421 APTTTKSAPTTTPKEPARTTKBAPTTPKEPARTTKBPTTPKEPARTTKBAPTTPK 480

```

```

QY 440 EPAPTAPKPADTTPKEPARTTPKBPAPTTTKPESTTPKEPARTTTKSAPTTTKBPAPT 499
Db 481 EPAPTAPKPADTTPKEPARTTPKBPAPTTTKPESTTPKEPARTTTKSAPTTTKBPAPT 540
QY 500 TTKSAPTTTKPESTTPKEPARTTPKBPAPTTTKPEPARTTPKBPAPTTTKBPAPT 559
Db 541 TTKSAPTTTKPESTTPKEPARTTPKBPAPTTTKPEPARTTPKBPAPTTTKBPAPT 600
QY 560 APTAPKEPARTTPKETAPTTPKKLTPTTPEKLAPTTPKEPARTTPPEELAPTTPEEPTPT 619
Db 601 APTAPKEPARTTPKETAPTTPKKLTPTTPEKLAPTTPKEPARTTPPEELAPTTPEEPTPT 660
QY 620 PEERAPTTPKAAAPNTTPKEPARTTPKBPAPTTTPKBPAPTTPKETAPTTPKGTAPTTLKEP 679
Db 661 PEERAPTTPKAAAPNTTPKEPARTTPKBPAPTTTPKBPAPTTPKETAPTTPKGTAPTTLKEP 720
QY 680 APTTPKAPKELAPTTTKETPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKBPAPTTPKG 739
Db 721 APTTPKAPKELAPTTTKETPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKBPAPTTPKG 780
QY 740 TAPTTLKEPARTTPKAPKELAPTTTKGPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKBPAPT 799
Db 781 TAPTTLKEPARTTPKAPKELAPTTTKGPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKBPAPT 840
QY 800 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 859
Db 841 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 900
QY 860 TKTPAATKPEMTTAKOKTTERDLRTTPTTAAKMTKETATTTTEKTESKITATTTQV 919
Db 901 TKTPAATKPEMTTAKOKTTERDLRTTPTTAAKMTKETATTTTEKTESKITATTTQV 960
QY 920 TSTTTQDTPPKITTLKTTTLAPKVTTKTITTEIMNKDEETAKPKDRATNSKATTPK 979
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTKTITTEIMNKDEETAKPKDRATNSKATTPK 1020
QY 980 POKETKAPKPTSTTKPKTWPVRVKPKTTPTPKOTSTMPELNPTSRIAEAMLOTTTRPN 1039
Db 1021 POKETKAPKPTSTTKPKTWPVRVKPKTTPTPKOTSTMPELNPTSRIAEAMLOTTTRPN 1080
QY 1040 QTNSKLVEVNPKSEDAGGAEGETPHMLRPHVFMPEVTDMYLPVNPQGIINPMLS 1099
Db 1081 QTNSKLVEVNPKSEDAGGAEGETPHMLRPHVFMPEVTDMYLPVNPQGIINPMLS 1140

RESULT 15
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseri, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 58:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1049 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-10-124-557-58
 Query Match 76.4%; Score 5561; DB 13; Length 1049;
 Best Local Similarity 95.5%; Pred. No. 2.2e-126;
 Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY	1	MAWKTLPIYLILLLSVFVIQOVSSQELSCKRCFESFERGECDCDAQCKKYDKCCPDYE	60
Db	1	MAWKTLPIYLILLLSVFVIQOVSSQELSCKRCFESFERGECDCDAQCKKYDKCCPDYE	60
QY	61	SFCAEVHNPTGPPSSKAPPPSGASQTIKSTKRSPPNKKTKKVIESEIEEHSVS	120
Db	61	SFCA-----EHSVS	70
QY	121	ENQSSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKKRTRKKKTPKPPVDE	180
Db	71	ENQSSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKKRTRKKKTPKPPVDE	130
QY	181	AGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKE	240
Db	131	AGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSIVNKE	190
QY	241	TTVETKETITNNKQSTDGKEKTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKG	300
Db	191	TTVETKETITNNKQSTDGKEKTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKG	250
QY	301	PALTTPKGPTPTTPKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTT	360
Db	251	PALTTPKGPTPTTPKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTT	310
QY	361	KEPAPTTKEPAPTTTKAPATTTSAPTTKEPAPTTPKAPATTTPKEPAPTTPKETPT	420
Db	311	KEPAPTTKEPAPTTTKAPATTTSAPTTKEPAPTTPKAPATTTPKEPAPTTPKETPT	370
QY	421	TTTPKEPAPTTKEPAPTTPKAPATTTPKEPAPTTPKAPATTTPKEPAPTTPKETPT	480
Db	371	TTTPKEPAPTTKEPAPTTPKAPATTTPKEPAPTTPKAPATTTPKEPAPTTPKETPT	430
QY	481	PAPTTTKSAPTTTKAPATTTSAPTTKEPAPTTTKAPATTTPKEPAPTTPKAPATTTP	540
Db	431	PAPTTTKSAPTTTKAPATTTSAPTTKEPAPTTTKAPATTTPKEPAPTTPKAPATTTP	490
QY	541	KEPAPTTKEPAPTTTKAPATTTSAPTTKEPAPTTTKAPATTTPKEPAPTTPKAPATTTP	600
Db	491	KEPAPTTKEPAPTTTKAPATTTSAPTTKEPAPTTTKAPATTTPKEPAPTTPKAPATTTP	550
QY	601	PTTPELAPTTPEPTPTTPKEPAPTTPKAAAPNTKEPAPTTPKAPATTTPKEPAPTT	660
Db	551	PTTPELAPTTPEPTPTTPKEPAPTTPKAAAPNTKEPAPTTPKAPATTTPKEPAPTT	610

Search completed: October 13, 2004, 11:53:15
 Job time : 145.669 secs

QY	661	KETAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKBPTSTSTSDKPAPTTKGTAPTT	720
Db	611	KETAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKBPTSTSTSDKPAPTTKGTAPTT	670
QY	721	PKAPATTTPKEPAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKPAPT	780
Db	671	PKAPATTTPKEPAPTTPKGTAPTTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKPAPT	730
QY	781	TPKETAPTTPKAPATTTPKAPATTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS	840
Db	731	TPKETAPTTPKAPATTTPKAPATTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELS	790
QY	841	AEPTPKALENSPKPEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPTTTAAPKMTKET	900
Db	791	AEPTPKALENSPKPEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPTTTAAPKMTKET	850
QY	901	ATTEKTTESKIATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTTKKIIITTEIMNKP	960
Db	851	ATTEKTTESKIATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTTKKIIITTEIMNKP	910
QY	961	EETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPTMPRVKPKTTTPRKMSTMPKE	1020
Db	911	EETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPTMPRVKPKTTTPRKMSTMPKE	970
QY	1021	LNPTSRIAEAMLQTTTTPNQTNPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD	1080
Db	971	LNPTSRIAEAMLQTTTTPNQTNPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD	1030
QY	1081	MDYLPVNPNOGIIINPMLS 1099	
Db	1031	MDYLPVNPNOGIIINPMLS 1049	

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:      October 13, 2004, 11:23:49 ; Search time 149.862 Seconds
              (without alignments)
              5233.063 Million cell updates/sec

Title:      SEQ1-F
Perfect score: 7276
Sequence:    1 MAKWTLPIYLLLLSVFIQ.....ARAITTRSGQTLSSKWNWNC 1363

```

```

Scoring table:  BLOSUM62
                  Gapop 10.0 , Gapext 0.1

Searched:      1825181 seqs, 575374546 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :      UniProt_Q2:*
            1:  uniprot_sprot:*
            2:  uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Query %	Score	Match	Length	DB	ID	Description
1	7261.9	99.8	1404	2	Q92954	Q92954	homo sapien
2	7232.9	99.4	1404	2	Q9BX49	Q9BX49	homo sapien
3	4852	66.7	933	2	Q62MZ5	Q62MZ5	homo sapien
4	4852	66.7	933	2	BAD18580	BAD18580	homo sapien
5	3886.9	53.4	1054	2	Q9UM99	Q9UM99	mus musculus
6	1714.7	23.6	401	2	Q77765	Q77765	bos taurus
7	1384.6	19.0	1225	2	Q9VR49	Q9VR49	homo sapien
8	1371	18.8	1719	1	MUC2_HUMAN	MUC2_HUMAN	homo sapien
9	1349.9	18.6	1761	2	Q7KTF6	Q7KTF6	homo sapien
10	1349.9	18.6	1761	2	Q6A64673	Q6A64673	homo sapien
11	1333	18.3	251	2	Q6DNM4	Q6DNM4	homo sapien
12	1324.3	18.2	3150	2	Q7PMD5	Q7PMD5	anopheles g
13	1237.4	17.0	3409	2	Q6SSB6	Q6SSB6	chlamydomon
14	1237.4	17.0	3409	2	QAS07044	QAS07044	chlamydomon
15	1130.4	16.4	1664	1	SLP1_CLOTM	SLP1_CLOTM	chlamydomon
16	1183	16.3	251	2	Q8QJF3	Q8QJF3	clostridium
17	1183	16.3	251	2	QAS83977	QAS83977	bos taurus
18	1172.4	16.1	3889	2	Q6SSB8	Q6SSB8	bos taurus
19	1172.4	16.1	3889	2	QAS07042	QAS07042	chlamydomon
20	1154.1	15.9	9234	2	Q7KTP5	Q7KTP5	chlamydomon
21	1154.1	15.9	9234	2	QAN10531	QAN10531	chlamydomon
22	1129.6	15.5	3432	2	Q8IR51	Q8IR51	chlamydomon
23	1129.2	15.5	3458	2	Q8IR52	Q8IR52	chlamydomon
24	1117.1	15.4	1349	2	Q8WU04	Q8WU04	chlamydomon
25	1109.2	15.2	5703	1	MU5B_HUMAN	MU5B_HUMAN	chlamydomon
26	1106.6	15.2	1795	2	Q76894	Q76894	chlamydomon
27	1105.9	15.2	1079	2	Q9N457	Q9N457	chlamydomon
28	1103.2	15.2	23015	2	Q8IQ18	Q8IQ18	chlamydomon
29	1103.2	15.2	23015	2	QAN10358	QAN10358	chlamydomon
30	1096.9	15.1	34350	2	Q8WZ42	Q8WZ42	chlamydomon
31	1069.7	14.7	2284	2	Q9VPG1	Q9VPG1	chlamydomon

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1

Q22954

ID Q22954 PRELIMINARY; PRT; 1404 AA.

AC Q22954;

DT 01-FEB-1997 (TREMBlrel. 02, Created)

DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE Megakaryocyte stimulating factor.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RN SEQUENCE FROM N.A.

RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,

RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,

RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,

RA Bhattacha S., Kriz R., Hewick R., Clark S.C.;

RT "Purification, Biochemical Characterization , and Cloning of a Novel

RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony

RT Stimulating Activity.";

RT Blood 78:279-279 (1991).

RN [2]

RN SEQUENCE FROM N.A.

RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,

RA Fitzgerald M., Scattreto J., Kelleher K., Preissner K., Kriz R.,

RA Jacobs K., Turner K.;

RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";

RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,

RL Mosher D.F. (eds.);

RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier

RL Science Publishers B.V. (1993).

RN [3]

RN SEQUENCE FROM N.A.

RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,

RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,

RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,

RA Bhattacha S., Kriz R., Hewick R., Clark S.C.;

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

RL EMBL; U70136; AAB09089.1; --.

DR HSSP; P04004; LOC0.

DR Genew; HGNC:9364; PRG4.

DR Go; GO:000283; P:cell proliferation; TAS.

DR InterPro; IPR000585; Hemopexin.

DR InterPro; IPR001212; Somatomedin_B.

DR Pfam; PF00045; Hemopexin; 2.

DR Pfam; PF01033; Somatomedin B; 2.

DR PRINTS; PR00022; SOMATOMEDINB.

DR SMART; SM00120; HX; 2.

DR SMART; SM00201; SO; 2.

DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.

DR PROSITE; PS00524; SOMATOMEDIN B; 2.

DR PROSITE; PS00524; SOMATOMEDIN B; 2.

DR SEQUENCE 1404 AA; 151090 MW; AARD7ADT9R1SP4GE CDS64.

SQL

Query Match		99.8%;	Score 7261.9;	DB 2;	Length 1404;	
Best Local Similarity		97.1%;	Pred. No. 2.7e-141;			
Matches 1363;		Conservative	0;	Mismatches	0;	Indels 41; Gaps 1;
QY	1	MAWKLPYVLLLLSVFVIQQVSSQ	-----	-----	-----	25
DB	1	MAWKLPYVLLLLSVFVIQQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF	60			
QY	26	-----ELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP	79			
DB	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP	120			
QY	80	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSHSVSENOESSSSSSSTIW	139			
DB	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSHSVSENOESSSSSSSTIW	180			
QY	140	KIKSSKNSAANRELQKLVKONKNKRTKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST	199			
DB	181	KIKSSKNSAANRELQKLVKONKNKRTKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST	240			
QY	200	TOHNKYSTSPKIITAKPINRPSLPNNSDTSKETSITVKNKETTVEKETTNNKQISTDG	259			
DB	241	TOHNKYSTSPKIITAKPINRPSLPNNSDTSKETSITVKNKETTVEKETTNNKQISTDG	300			
QY	260	KEKTSKETSQTSIEKTSKADLPTSKVLAKPTPKAETTTKGPALTTPEKPTTPKEPAS	319			
DB	301	KEKTSKETSQTSIEKTSKADLPTSKVLAKPTPKAETTTKGPALTTPEKPTTPKEPAS	360			
QY	320	TTPEKPTTPKSAPTTPKBPATTTTKSAPTTKBPATTTKBPATTTKBPATTTKBP	379			
DB	361	TTPEKPTTPKSAPTTPKBPATTTTKSAPTTKBPATTTKBPATTTKBPATTTKBP	420			
QY	380	APTTPKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	439			
DB	421	APTTPKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	480			
QY	440	EPAPTAKKAPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	499			
DB	481	EPAPTAKKAPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	540			
QY	500	TTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	559			
DB	541	TTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	600			
QY	560	APTAPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	619			
DB	601	APTAPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	660			
QY	620	PEEPAPTTKAAAPNTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	679			
DB	661	PEEPAPTTKAAAPNTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	720			
QY	680	APTTPKBPAPKELAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	739			
DB	721	APTTPKBPAPKELAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	780			
QY	740	TAPTTLKBPATTTKBPAPKELAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	799			
DB	781	TAPTTLKBPATTTKBPAPKELAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	840			
QY	800	KPAPTTPPTTSEVSTPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	859			
DB	841	KPAPTTPPTTSEVSTPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	900			
QY	860	TKTBAATKEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETAITTEKTTESKITATTQV	919			
DB	901	TKTBAATKEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETAITTEKTTESKITATTQV	960			
QY	920	TSTTTQDTPPKITLTKTTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK	979			
DB	961	TSTTTQDTPPKITLTKTTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK	1020			
QY	980	PQKPTKAPKKPTSTKKPKTMPVRVKPTTTPPKMTSTWPELNPTSRIAEAMLQTTTREN	1039			
Query Match		99.4%;	Score 7232.9;	DB 2;	Length 1404;	
Best Local Similarity		96.8%;	Pred. No. 1.1e-140;			
Matches 1359;		Conservative	0;	Mismatches	4;	Indels 41; Gaps 1;
QY	1	MAWKLPYVLLLLSVFVIQQVSSQ	-----	-----	-----	25
DB	1	MAWKLPYVLLLLSVFVIQQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF	60			
QY	26	-----ELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP	79			
DB	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP	120			
QY	80	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSHSVSENOESSSSSSSTIW	139			
DB	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSHSVSENOESSSSSSSTIW	180			

```

QY 140 KIKSSKNGAANRELQKLVKONKKNRKKKPTPKPPVDEAGSLDNGDFKVTTPDTGT 199
Db 181 KIKSSKNGAANRELQKLVKONKKNRKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240
QY 200 TOHNKVSPTSPIITAKPNINRSLPNSTSKTSLSITVNKETVETKETTNTKQSTSDG 259
Db 241 TOHNKVSPTSPIITAKPNINRSLPNSTSKTSLSITVNKETVETKETTNTKQSTSDG 300
QY 260 KEKTSKETSOSIEKTSADLAPTQKVLAKPTPKAETTTKGPALTTKPEPTTTTKEPAS 319
Db 301 KEKTSKETSOSIEKTSADLAPTQKVLAKPTPKAETTTKGPALTTKPEPTTTTKEPAS 360
QY 320 TTPKEPTTTTKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 379
Db 361 TTPKEPTTTTKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 420
QY 380 APTTTKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 439
Db 421 APTTTKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 480
QY 440 EPAPTAPKAPAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 499
Db 481 EPAPTAPKAPAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 540
QY 500 TTKSAPTTPKBPAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 559
Db 541 TTKSAPTTPKBPAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 600
QY 560 APTAPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 619
Db 601 APTAPKBPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 660
QY 620 PEPAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 679
Db 661 PEPAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
QY 680 APTTPKBPAPKELAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 739
Db 721 APTTPKBPAPKELAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 780
QY 740 TAPTTLKEAPTTPKBPAPKELAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 799
Db 781 TAPTTLKEAPTTPKBPAPKELAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 840
QY 800 KPAPTTPPTPTTSEVSTPTTKEPTTIHKSDESTPELSABPTPKALENSKPEGVPT 859
Db 841 KPAPTTPPTPTTSEVSTPTTKEPTTIHKSDESTPELSABPTPKALENSKPEGVPT 900
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKPMKETAATTTBKTTESKITATTTOV 919
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKPMKETAATTTBKTTESKITATTTOV 960
QY 920 TSTTTQDTTTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEBETAKPKRATNSKATTPK 979
Db 961 TSTTTQDTTTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEBETAKPKRATNSKATTPK 1020
QY 980 POKPTKAPKPTSTTKPKTTPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRN 1039
Db 1021 POKPTKAPKPTSTTKPKTTPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRN 1080
QY 1040 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVPTDMDYLPRVFNQGIINPMLS 1099
Db 1081 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVPTDMDYLPRVFNQGIINPMLS 1140
QY 1100 DETNINCNGKPVGLTTLRNGTLVARGHYFWMLSPPSPSPARRITEVWGIPSPIDTFT 1159
Db 1141 DETNINCNGKPVGLTTLRNGTLVARGHYFWMLSPPSPSPARRITEVWGIPSPIDTFT 1200
QY 1160 RCNCEGKTFKFDQSVWFRTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPSVY 1219
Db 1201 RCNCEGKTFKFDQSVWFRTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPSVY 1260

```

```

QY 1220 FFKRGSGIOQYIKQEPVQKCPGRPALNYPYVGEMTQVRRRFFERAIGPSQHTTIRIQY 1279
Db 1261 FFKRGSGIOQYIKQEPVQKCPGRPALNYPYVGEMTQVRRRFFERAIGPSQHTTIRIQY 1320
QY 1280 SPARLAYQDKGLVHNEVKVSLIWRGLPNVVTSAISLNPRIKPDGDYDYAFSKDQYNNIDV 1339
Db 1321 SPARLAYQDKGLVHNEVKVSLIWRGLPNVVTSAISLNPRIKPDGDYDYAFSKDQYNNIDV 1380
QY 1340 PSRTARAITTRSGQTLKSVWYNCP 1363
Db 1381 PSRTARAITTRSGQTLKSVWYNCP 1404

ID Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Irie R., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wgatusuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131434; BAD18580.1; -;
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 1.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 66.7%; Score 4852; DB 2; Length 933;
Best Local Similarity 68.3%; Pred. No. 6.1e-92;
Matches 931; Conservative 0; Mismatches 2; Indels 430; Gaps 1;

QY 1 MAWKTLPIYLLLLLSVFIQVSSQELSCKRCFFSFERGRCDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLLSVFIQVSSQELSCKRCFFSFERGRCDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNPSTPPSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEITBEHSVS 120
Db 61 SFCAEVHNPSTPPSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEITBEHSVS 120
QY 121 ENQESSSSSSSSSSSTTKIKSSKNANRELQKLVKONKKNRTKKKTPKPPVDE 180
Db 121 ENQESSSSSSSSSSSTTKIKSSKNANRELQKLVKONKKNRTKKKTPKPPVDE 180
QY 181 AGSGLDNGDFKVTTPDTSTTOHNKVSPTSPIITAKPNINRSLPNSTSKTSLSITVNKE 240
Db 181 AGSGLDNGDFKVTTPDTSTTOHNKVSPTSPIITAKPNINRSLPNSTSKTSLSITVNKE 240
QY 241 TVTETKETTNTKQSTSDGKEKTTSAKETQSTIEKTSADLAPTQKVLAKPTPKAETTTTKE 300
Db 241 TVTETKETTNTKQSTSDGKEKTTSAKETQSTIEKTSADLAPTQKVLAKPTPKAETTTTKE 300
QY 301 PALITPKPTTTPKEPASTTPKEPTPTTIKSAPTTPKBPAPTTPKAPTTPKAPTTPK 360

```

Db	301		PALTTTKEPTTTPKEPASTTKEPTTIIKSAPTTTPKEPAPTTTTSAPTTTKEPAPTTT	360
Qy	361		KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	420
Db	361		KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	370
Qy	421		TTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	480
Db	371	-----	-----	370
Qy	481		PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	540
Db	371	-----	-----	370
Qy	541		KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	600
Db	371	-----	-----	370
Qy	601		PTTPEELAPTTTPEEPTTPEEAPTTTAAAPNTPEKAPTTTKEPAPTTTKEPAPTTT	660
Db	371	-----	-----	370
Qy	661		KETAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	720
Db	371	-----	-----	370
Qy	721		PKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	780
Db	371	-----	-----	370
Qy	781		TPKETAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	840
Db	371	-----	-----	370
Qy	841		AEPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	900
Db	411		AEPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	470
Qy	901		ATTTEKTESKIATTTTQVSTTTTQDTTTPFKITLKTITLAPKVTITTKTITTTTINMKP	960
Db	471		ATTTEKTESKIATTTTQVSTTTTQDTTTPFKITLKTITLAPKVTITTKTITTTTINMKP	530
Qy	961		BETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKPTMVRPKPTTTPPKMTSTMP	1020
Db	531		EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKPTMVRPKPTTTPPKMTSTMP	590
Qy	1021		LNPTSRIAEAMLOTTTTRNQTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMDEVTPD	1080
Db	591		LNPTSRIAEAMLOTTTTRNQTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMDEVTPD	650
Qy	1081		MDYLPVFNQGIINPMLSDETNICNGKPVVDGLTLRNGTLVAFRGHYFWMLSPPSPSP	1140
Db	651		MDYLPVFNQGIINPMLSDETNICNGKPVVDGLTLRNGTLVAFRGHYFWMLSPPSPSP	710
Qy	1141		ARRITEVWGIPSIDTFTFRNCCEKTPFFKDSQYWRFTNDIKDAGYKPIFKGGGLTG	1200
Db	711		ARRITEVWGIPSIDTFTFRNCCEKTPFFKDSQYWRFTNDIKDAGYKPIFKGGGLTG	770
Qy	1201		QIVAIATKAYKNWPSVYFFKRGSIQYIYKQBPVKQCGRRPALNVPVYGETQVRR	1260
Db	771		QIVAIATKAYKNWPSVYFFKRGSIQYIYKQBPVKQCGRRPALNVPVYGETQVRR	830
Qy	1261		RRFERAIGPSQHTTIRIQSPARLAYQDKGVHLNVEKVSILWRGLPNVVTSAISLPIRK	1320
Db	831		RRFERAIGPSQHTTIRIQSPARLAYQDKGVHLNVEKVSILWRGLPNVVTSAISLPIRK	890
Qy	1321		PDGYDYAFSKQOYINIDVPSRTARAITTTSQGLTSKWNVNC	1363
Db	891		PDGYDYAFSKQOYINIDVPSRTARAITTTSQGLTSKWNVNC	933

790	QY	QY	PKBAPPTPKKPAFTTETPPPTTSEVSTPTTTTKEPTTIHKGDDESTPBLSAEPTPKALE	849
602	Db	Db	PKPEPTTPKKPEPTT	617
850	QY	QY	NSPKEPGVPTTKTGAATKDEMTTAKDKITRDLRTTTPETTAAPKMTKETATTTTEKITTE	909
			:	
618	Db	Db	-----TSPKT-----	622
910	QY	QY	SKITATTTQVTSITTTQDTPPKITLTKTTLAPKVTITTKTITTTIEMNKPETAKPKDR	969
623	Db	Db	-----TTLKATTLAPKVAPAE-----BIQNKPEITTPASED	654
970	QY	QY	ATNSKAT-----TPKP-QKPTKAPKPTSTCKPMTPRVAKPXTTTPRMTSTWPE	1029
			:	
655	Db	Db	SDDSKTTLKPKQPKAPKPTKKPKAPKPTSTKKPKT-PKTRKPKTTPAPLKTTTSATPE	713
1021	QY	QY	LNPTSRISAEAMLQTTTRNQTNPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD	1088
714	Db	Db	LNMTB--LEVMLPTTTPKQTPNPETAENVPDHEDAGGEGEKP-LIPGPPVLLFPFPAIPG	770
1081	QY	QY	MDYLPVRVENQGIINPMLSDSTNICNGKVPDVLTLRNGTLVAFRGHYFWMLSPFPPSP	1144
771	Db	Db	TDLLAGRLNRGININPMSDETNICNGKVPDVLTLRNGTLVAFRGHYFWMLNPPFPSP	830
1141	QY	QY	ABRIEIVWGISPIDVTFRNCCEKTPFFKDSQYWRFTNDIKDAGYPKP1EKGFGGLTG	1200
831	Db	Db	PRRIEIVWGISPIDVTFRNCCEKTIFFKDSQYWRFTNDVDPGYPKQ1VKGFGGLTG	890
1201	QY	QY	QTVAA1STAKYKNWPESVYFFKRGSSIQQY1YKQBPVQKCGRRPALNYPVYGEMTQVRR	1266
891	Db	Db	KIVAALSIAKYKDRPESVYFFKRGSGNIQQYTYKQBPMMKCTGRRPALNYSVYGEEAQVRR	950
1261	QY	QY	RRPERAIGSQTHTRIQV5-PARLAYQDKGLVHNEVKV5L1AREGLBNVYTSALSIPNR	1311
951	Db	Db	RRFERAVGPFQTHTRIRHYSVPMRV5YQDKGLHNEVKV5TMMRGFNVVTSALTIPNR	1011
1320	QY	QY	KPDGYDYAFSKDOYVNDVPSRTARALTTSGOTLSKWNCP	1363
1011	Db	Db	KPDGYDYAFSKDOYVNDVPRTRARALTTSGOTLSKIWNCP	1054

```

Q9VR49  RESULT 7
ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgsl; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI TaxID=7227;
[1]_
SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

```

Q9VR49  RESULT 7
ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
CG 3047-PA.
DE Name=Sgsl; ORFNames=CG3047;
GN Drosophila melanogaster (Fruit fly).
OS Drosophila melanogaster; Hexapoda; Insecta; Pterygota;
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI TaxID=7227;
[1]_TaxID=7227;
SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

Query Match

Best Local Similarity 25.1%; Pred. No. 3.2e-20;	
Matches 378; Conservative 135; Mismatches 428; Indels 564; Gaps 59;	
QY	10 LLLLLSVFLQQVS-----SQEL--SKGRCFESFGRGECDDACKKYD 53
Db	6 IFVTVILIIQVKNVANDWMDQGGSEEIIPCGGGTIYVDPVQPCDTS----- 59
QY	54 KCCPDYESFCAEVHNPTGPPSSKAPPPSGASQTIKSTTKESPKNKKTKVLESEI 113
Db	60 -----NPTTKPKQTKRP-----KSTRRTTKTKRP--RRKTK- 92
QY	114 TBEHVSVENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKNRTKKP-T 172
Db	93 -----WTKRA-----TKRTTKRTTRRPPT 113
QY	173 KPPVVDBAGSLDNGDFKVTTPDT-----STIQNKVS-----TSPKTTA 214
Db	114 PKTP-----DTTDPITITGABCTCDRTTASSTDTSTDRITVNTDWTPLCTDT 164
QY	215 KPINRPPLPPNSDTSKETSITVNKEITVE-----TKETTTNKQSTDGEKT- 263
Db	165 PPCT-----CSESSNAIPSSPCIDTSVITPSPCTQETITTFPCSTQGIQITP 214
QY	264 --TSAKETQSIEKTSAKDLAPTSKVLKPTPKAET--TKGPALTPKEPTPTTPKEPAST 320
Db	215 CTCAOTT-----TTPRSTTTTTSRPTTTTPRSTTTTTSRPTTT 254
QY	321 TPKEPTPTTIKSAPTPKEPAPTTTKAP--TTPKEPAPTTTKEPAPTPKEPAPTTTKE 378
Db	255 TPRSITTTTTRPTTTTPRCWTITSTCAPTTTTPRSTTTTTSRPTTTTPRSTTTTSCS 314
QY	379 PAPTTTKSAPT--PKEPAPTP-----KKPAPTPKE-----PAPTPKE- 417
Db	315 PTRTPRSTTTTTSRPTTTTPRCWTITTPSTTCIRPPTTTTPRSTTTTCTSGPTTTTPRST 374
QY	418 -----PTPTTPKEPAPT--TKEPAPTPKEPAPTAPKAPTPKE-----PAPT 461
Db	375 TTCTSGPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTT 434
QY	462 PKEPAPTTTKEPSPPTTKEPAPTTTKSAPTTTKEPAPTTTKAP--TTPKEPSPPTTKP 519
Db	435 PRSITTTTSGPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTSGF 494
QY	520 APTTPKE-----PAPTPPK-----PAPTPPK-----PAPTPPKPAPT 555
Db	495 TTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTT 554
QY	556 TKKAPAPKE-----PAPTPKEPAPTTPKLTPTTPKELAPTTPEKAPTTPEL 607
Db	555 TSGPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRST 614
QY	608 APTTPPEPTTPEE-----PAPTTPKAAAPNTPKEPAPTPE-----PAPT 651
Db	615 TTISTSGPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTT 674
QY	652 PKB-----PAPTPKBT-----APTPKGTAPTTLKEPAPTTPKP----- 687
Db	675 PRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPT 734
QY	688 ---APKELAPTTTKBPTST--SDKPAPTPKGTAPTTPEKAPTPKE----- 731
Db	735 TTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTT 794
QY	732 ---PAPTPKGTAPTTLKEPAPTTPKK-----PAPKELAPTTKGTPTST-- 773
Db	795 TSGPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRST 854
QY	774 ---SDKPAPTPKGTAPTTPEKAPTTPK-----PAPTPPET-----PPTT 813
Db	855 TTSTTSCPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTSRPTTTT 914
QY	814 SEVSPTTTTKEPTTHKSPDSTSELSAEPPTKALENSPKEPVGPTTKTFAATKPEMTT 873

Db	915	FRSTTTCTSGTT--TTPTSTTKTSTCAPT-----TTTPTSTTTTSTSRPTTTTPTSTTT	968	RL	J. Clin. Invest. 88:1005-1013(1991).
Qy	874	AKDKTTERDLRTTPETTAAPKWKETATTTKTESKITATTQVTSVTTQDTPPFKIT	933	CC	- FUNCTION: Coats the epithelia of the intestines, airways, and other mucus membrane-containing organs. Thought to provide a protective, lubricating barrier against particles and infectious agents at mucosal surfaces.
Db	969	T--TTSRPTTTTPRSTT--TPSTSRPTTTTPRSTT-----TTSRPTTTTTPRSTT	1015	CC	- SUBUNIT: Multimeric.
Qy	934	TLKTTTLAPKVTTKTKITTTTTEIMNKPEETAKPKDRATNSKATTPKQKTPKAPK	988	CC	- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors, bronchus, cervix and gall bladder.
Db	1016	--KTSICAPTITTPRSTTTT-----TSRPTTTTPRSTTTTTSRPTTTTTPRSTTTP	1065	CC	- PTM: All cysteine residues are involved in intrachain or interchain disulfide bonds (By similarity).
Qy	989	---KPTSTKPKTMRVRKPKTPPKMT-----STMPENPTSRIAEAMLOTTTR	1037	CC	- POLYMORPHISM: The number of repeats is highly polymorphic and varies among different alleles.
Db	1066	CTSRPTTTTPRSTTTTTPRSTTTTPRSTTTPCPTTPASPR-----TTTTRPCP	1118	CC	- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
Qy	1038	-----PNOT-----PNSKLVEVNPKSEDAGAGETPHMLLRPHV	1072	CC	- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
Db	1119	CHPQPPYQIYPPMSWYFNTYFNPVWPQFN-----	1149	CC	- SIMILARITY: Contains 2 VWFC domains.
Qy	1073	FMEVTPDMYLPVRNQGIIINPMLSDETNICNGKXPVDGLTLRNGTLVAERGHVWML	1132	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
Db	1150	---PVVPQWQMPGPNQ-----	1167	CC	EMBL; L21998; AAB95295.1; -
Qy	1133	SPESP---SPARRITEVMGIPSPIDIVFTRCNCCEGKTFPKDSQVWRFTNDKAGYK	1189	DR	EMBL; M74027; AAA59875.1; -
Db	1168	LPGYPOOLPPP-----LPSP-----	1190	DR	EMBL; M94131; AAA59163.1; -
Qy	1190	P-----IPKFGGGLTGQIYAALSTAKYKNWPSVYFKRGGSIQOYIYKOE	1235	DR	EMBL; M94132; AAA59164.1; -
Db	1191	PPVVPQGDCCNICENLLKGVK-----	1215	DR	PIR; A49963; A43932.
Qy	1236	PVQKC 1240		DR	Genew; HGNC:7512; MUC2.
Db	1216	LIRRC 1220		DR	MIM; 158370; -
RESULT 8				DR	InterPro; IPR002919; Cysrich TIL.
MUC2 HUMAN				DR	InterPro; IPR006208; Cys_knotE.
ID MUC2 HUMAN STANDARD; PRT; 5179 AA.				DR	InterPro; IPR006207; Cys_knot_C.
AC Q02817; Q14878;				DR	InterPro; IPR006209; EGF_like.
DT 01-JUN-1994 (Rel. 29, Created)				DR	InterPro; IPR001007; VWFC.
DT 01-NOV-1997 (Rel. 35, Last sequence update)				DR	InterPro; IPR001846; VWFC_D.
DT 05-JUL-2004 (Rel. 44, Last annotation update)				DR	Pfam; PF00007; Cys_knot; 1.
DE Mucin 2 precursor (intestinal mucin 2).				DR	Pfam; PF01826; TIL; 1.
GN Name=MUC2; Synonyms=SMUC2;				DR	Pfam; PF00093; VWC; 1.
OS Homo sapiens (Human).				DR	Pfam; PF00094; VWC; 4.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				DR	SMART; SM00041; CT; 1.
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				DR	SMART; SM00214; VWC; 2.
OX NCBI_TaxID=9606;				DR	SMART; SM00216; VWD; 4.
RN [1]				DR	PROSITE; PS01185; CTCK_1; 1.
RP SEQUENCE FROM N.A.				DR	PROSITE; PS01225; CTCK_2; 1.
RC TISSUE=Intestine;				DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.
RX MEDLINE=94132002; PubMed=8300571;				DR	PROSITE; PS01208; VWFC_1; 2.
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				DR	PROSITE; PS0184; VWFC_2; 2.
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.				KW	Glycoprotein; Repeat; Signal.
RT Identification of the amino terminus and overall sequence similarity				FT	Potential.
RT to prepro-von Willebrand factor."				FT	SIGNAL 1 20
RL J. Biol. Chem. 269:2440-2446(1994).				FT	CHAIN 21 5179
RN [2]				FT	Mucin 2.
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				FT	DOMAIN 1401 1747
RC TISSUE=Colon;				FT	REPEAT 1401 1416
RX MEDLINE=93016075; PubMed=1400449;				FT	REPEAT 1417 1432
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				FT	REPEAT 1433 1448
RA Kim Y.S.;				FT	REPEAT 1449 1464
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located				FT	REPEAT 1465 1471
RT both upstream and downstream of its central repetitive region."				FT	REPEAT 1472 1478
RL J. Biol. Chem. 267:21375-21383(1992).				FT	REPEAT 1479 1494
RN [3]				FT	REPEAT 1495 1517
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				FT	REPEAT 1518 1533
RX MEDLINE=91358717; PubMed=1885763;				FT	REPEAT 1534 1556
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				FT	REPEAT 1557 1572
RA Petersen G.M., Kim Y.S.;				FT	REPEAT 1573 1596
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays				FT	REPEAT 1597 1612
RT and polymorphism."				FT	REPEAT 1613 1635
				FT	REPEAT 1636 1651
				FT	REPEAT 1652 1675
				FT	REPEAT 1676 1683
				FT	REPEAT 1684 1699
				FT	REPEAT 1700 1715

FT	REPEAT	1716	1731	15.
FT	REPEAT	1732	1747	16.
FT	DOMAIN	4815	4886	VMFC 1.
FT	DOMAIN	4924	4991	VMFC 2.
FT	DOMAIN	5075	5160	CTCK.
FT	DISULFID	5075	5122	By similarity.
FT	DISULFID	5089	5136	By similarity.
FT	DISULFID	5098	5152	By similarity.
FT	DISULFID	5102	5154	By similarity.
FT	DISULFID	?	5159	By similarity.
FT	CARBOHYD	163	163	N-linked (GlcNAc. .)
FT	CARBOHYD	423	423	N-linked (GlcNAc. .)
FT	CARBOHYD	670	670	N-linked (GlcNAc. .)
FT	CARBOHYD	770	770	N-linked (GlcNAc. .)
FT	CARBOHYD	894	894	N-linked (GlcNAc. .)
FT	CARBOHYD	1139	1139	N-linked (GlcNAc. .)
FT	CARBOHYD	1154	1154	N-linked (GlcNAc. .)
FT	CARBOHYD	1215	1215	N-linked (GlcNAc. .)
FT	CARBOHYD	1230	1230	N-linked (GlcNAc. .)
FT	CARBOHYD	1246	1246	N-linked (GlcNAc. .)
FT	CARBOHYD	1787	1787	N-linked (GlcNAc. .)
FT	CARBOHYD	1820	1820	N-linked (GlcNAc. .)
FT	CARBOHYD	4339	4339	N-linked (GlcNAc. .)
FT	CARBOHYD	4351	4351	N-linked (GlcNAc. .)
FT	CARBOHYD	4362	4362	N-linked (GlcNAc. .)
FT	CARBOHYD	4373	4373	N-linked (GlcNAc. .)
FT	CARBOHYD	4422	4422	N-linked (GlcNAc. .)
FT	CARBOHYD	4438	4438	N-linked (GlcNAc. .)
FT	CARBOHYD	4502	4502	N-linked (GlcNAc. .)
FT	CARBOHYD	4616	4616	N-linked (GlcNAc. .)
FT	CARBOHYD	4627	4627	N-linked (GlcNAc. .)
FT	CARBOHYD	4752	4752	N-linked (GlcNAc. .)
FT	CARBOHYD	4787	4787	N-linked (GlcNAc. .)
FT	CARBOHYD	4881	4881	N-linked (GlcNAc. .)
FT	CARBOHYD	4888	4888	N-linked (GlcNAc. .)
FT	CARBOHYD	4955	4955	N-linked (GlcNAc. .)
FT	CARBOHYD	4970	4970	N-linked (GlcNAc. .)
FT	CARBOHYD	5019	5019	N-linked (GlcNAc. .)
FT	CARBOHYD	5038	5038	N-linked (GlcNAc. .)
FT	CARBOHYD	5069	5069	N-linked (GlcNAc. .)
FT	CONFLICT	1351	1351	H -> L (in Ref. 3).
FT	CONFLICT	1412	1412	T -> S (in Ref. 3).
FT	CONFLICT	1449	1449	L -> P (in Ref. 3).
FT	CONFLICT	1504	1504	M -> T (in Ref. 3).
FT	CONFLICT	4192	4192	G -> S (in Ref. 2).
SQ	SEQUENCE	5179	5179	AA; 540295 MW; 85CD7571FB9A5663 CRC64;

Query Match			
Best Local Similarity 23.8%; Pred. No. 9.6e-19;			
Matches 443; Conservative 75; Mismatches 502; Indels 840; Gaps 69;			
QY	2	AWKTLPIY-----LALLSVFVI-----QQVSSQE 26	
Db	1020	SMKEATCPDVSTNPBPSLNPHRSWAERKQCSILKSSVFSI	CHSKVDKPFVEACVHDS 1079
QY	27	LSCK--GRFESPERGECDDA-----QCKYDKC-----CPDYESFCAEVHNT 70	
Db	1080	CSDTGGDC-----ECFSAVASVAQECTKEGACVFWRTDLCF---IFC-DYINPP 1127	
QY	71	SP-----PSSKAPP-----80	
Db	1128	HECEWHEPCGNRSFETCRTHINGHSNISVYLEGCVPRCPKDRPIVEBDLKCVCVTADKC 1187	
QY	81	-----PSGAS---OTIKS-----90	
Db	1188	GCVEDTHYPGASVTEETCKSCVCTNSQVVCVRPEEGKLNQTDGAFYCEICGPN 1247	
QY	91	-----TTKSPKPPNKKTKKVI SEBEITBEHVSSENQESSSS 129	
Db	1248	TVEKHFNICSIITRPSLTATFTITLPTPTSTTT-----TTT 1285	
QY	130	SSSSSSSTTWKTKSSKNKSANRELQKKLVKONKNRTKKKTPPKPVV-----DEAG 182	
Db		::: ::::	

Db	1286	TTTTSSIV-----LSTPKLCLLWSDWINEHPS 1315	
QY	183	SGLDNGDFK-----VTFPDISTTQH-NKVSSTSPK-----210	
Db	1316	SGSDDGDRPPDVGVCAPEDIECRSVKDPHLSEHQGVQCDVSVGFICKNEDQFGNP 1375	
QY	211	-----ITTAKPINPRSLPPNSDTSKETSLSLVNKEITVETKEITT 250	
Db	1376	FGLCVDYKIRVNCWPMDCIKTITPSPPPTTTTLLPPTTTPSPPTT-----TTT 1430	
QY	251	TNKOTSTDGKEKTTSAKETQSIKTSKDLAPTSKVLAKPTKARTTKGPAITTPKEPT 310	
Db	1431	TPPTTTTPSPPTTITTTT-----LPTT-----TPSEPITSTTTTPPTTTPSPPT 1474	
QY	311	-----PTTPKBPASTTPKEPTTIKSAK-TPKBP-----APTITKSAK-TPKBP 356	
Db	1475	TTSPPTTTTPSPPTTTTTPPTTTPSPPTTTPASTTTLPTTTPSPPTTITTTTP 1534	
QY	357	PTTTKEPAPTTPKBP-----APTITKEPAPTTPKSAK-TPKBPAPTTPKBPAPTTP 407	
Db	1535	PTTTTPSPPTTITPTPTSTTTLPPTTTPSPPTTTPPTTTPPTTTPSPPTTITTT 1594	
QY	408	KEPAPTTPKBPPTTPKBPAPTTPKBPAPTTPKBP-----APTAPKBPAPTTPKBPAPT 460	
Db	1595	TPPTTTTPSPPTTITTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1654	
QY	461	TPKBPAPTTPKBPPTTPKBPAPTTPKSAPTTTKEPAPTITTKSAK-TPKBPSPPTTK 517	
Db	1655	TTSPPTTTTPSPPTTITTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1714	
QY	518	EPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP-----572	
Db	1715	PSSTTTP-SPPTTTMTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1773	
QY	573	KETAPTTPKKLTP-----585	
Db	1774	STITPTP--CVPLCNWGLDSGKNFHKPGGDTLIGDVGPGFMAANISCRATMVPDV 1831	
QY	586	-----TPPEKLAPTTP 597	
Db	1832	PIGQLQTVVCDVSVGLICKNEQKPGGVIPMAFLNVEINVOCCECVQPTMTTITTE 1891	
QY	598	KPAPTTPPEELAPTTPPEPTP-----TPPEBPAPT-----626	
Db	1892	NPTPTTPTTITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1951	
QY	627	-----TPKAAAPNTP-----KEPAPTTPKBP-----APTTPKBPAPT-TPKETA 664	
Db	1952	TVPTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2011	
QY	665	PTTPKGTAPTTLKEPAPTTPKBPAPKELAPTTPKBP-----TSTTSKPA 709	
Db	2012	TTPTTITTTTTPPTPTPTGQTPTTPTTPTTITTTTTPPTPTPTGQTPTTPTTPTTPTTPT 2071	
QY	710	PT-----TPKGT-----APTTPKBPAPT-----726	
Db	2072	PTPTGQTPTTPTTPTTPTPTPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2131	
QY	727	-----TPKGT-APT-----LKEPAPT 752	
Db	2132	TTTITTTTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2191	
QY	753	PKKAPKELAPTTPKBP-----TSTTSKBPAP-----TPKETAPT 788	
Db	2192	TQPTTPTTPTTPTTPTPTGQTPTTPTTPTTITTTTTPPTPTPTGQTPTTPTTPTTPTTPT 2251	
QY	789	TPKBPAP-----TPKBPAPT-----TPETPP-----PT 812	
Db	2252	VPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2311	
QY	813	TSEVS-----TPTTKBPTTIHKSPDESTPESAEPTPKALENSKBP-----855	
Db	2312	TTPTTITTTTTPPTPTPTGQTPTTPTTPTTITTTTTPPTTPTTPTTPTTPTTPTTPTTPTTPT 2368	

Db	252	RKSSTA-----	-----KTTREPTTK-----	-----RETERTTKEPST	279
QY	198	S-TTOHNVSTSPKITTAKPINDRPSLPNSDTSKETSITVANKETV-----	-----	-----	243
Db	280	SKITTHE-----	-----TIAEPATKKITHIPT-----	-----TOKSTTLRIEETPRTRKSSSTARTREPT	329
QY	244	---ETKETT-----	-----TNKOTSDGKEKTSISAKETO-----	-----SIEKTSAKDIAPTS	284
Db	330	TKRETERTTQBPSTSKTTHETTABPAKKTHTHEPTTQKSTTLRIEETPRTRK-----	-----	-----SSTA	387
QY	285	KVLAKETPKAET-----	-----TTKGAULTTPKEPTPTPKEPAS-----	-----TTPKEPT-----	328
Db	388	KITREPTTKRETERTTKEP-----	-----STSKTTHETAAEPATKKTHETPTTQKSTTLRIEETPT	-----	445
QY	329	TIKSA-----	-----PTTPKE-----	-----PA-----	348
Db	446	TKSSSTARTREPTTKRETERTTQBPSTSKTTHETTABPAKKTHTHEPTTQKSTTLRI	-----	-----	505
QY	349	---PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP-----	-----APTTPKSA-----	-----PTTP	391
Db	506	TEEPTTRKSSATAKTRE-----	-----PTTKRETERTTQBPSTKTTTHETTAETAATKKTHTHEPTTQ	-----	563
QY	392	KEPA-----	-----PTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPA-----	-----PTTPKEPA-----	434
Db	564	KSTTLRIEETPRTRKSSSTARTIRE-----	-----PTTKRETERTTQBPSTKTTTHETTAETAATKKTHTHEPTTQ	-----	621
QY	435	---PTTPKEPA-----	-----PTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPT-----	-----TPK	479
Db	622	THEPTTQKSTTLRIEETPRTRKSSSTARTIRE-----	-----PTTKRETERTTQBPSTKTTTHETTAETAATKKTHTHEPTTQ	-----	679
QY	480	EPAPTTPKSAPTTPKEPA-----	-----PTTPKSA-----	-----PTTPKEPSPTTPKEPA-----	520
Db	680	EPATKKTHETPTTQKSTTLRIEETPRTRKSSATAKTREPTTKRETERTTKEPTTRKTTT	-----	-----	739
QY	521	-----	-----PTTPK-----	-----EPA-----	544
Db	740	IKTTEEPTKKTHETPTTKKSTTLKPTSEPTTRKSTTTKTREPTTKRKTTERTTKEPTT	-----	-----	799
QY	545	-----PTTPKEPAPTTPKKA-----	-----PTAPK-----	-----EPA-----	571
Db	800	RKTTHKTTEEPTINK-----	-----TTTKKTHETPTTKKSTTLKPTSEPTTRKSTTTKTREPTTKRKTTERTTKEPTT	-----	855
QY	572	PKETA-----	-----PTTPKKILT-----	-----PTT-----	595
Db	856	KRETERTTQBPSTSKTTHETTABPAKKTHTHEPTTQKSTTLRIEETPRTRKSSATAKT	-----	-----	915
QY	596	PEKPAPTTPBELAPTPEEPT-----	-----PTTPB-----	-----EPA-----	628
Db	916	RE-----	-----PTTKRETERTTKEPTTRKKTTHKTTEEPTTKKTHTHEPTTKKSTTLKPTSEPTTR	-----	972
QY	629	KAAAPNTKEPAPTTPKEPAPTTPKEPA-----	-----	-----PTTPKEPA-----	670
Db	973	KTSSTKTIRE-----	-----PTTKRKTTERTTKEPTTRKKTTHKTTEEPTTKKTTHKTTEEPTTKKTTHKTTEEPTTKKS	-----	1030
QY	671	TA-----	-----PTTLKEPAPTTPKKAPKELA-----	-----PTTKKPTSTTSDDKA-----	710
Db	1031	TTLKPTSEPTTRKSTTTKTREPTTKETVTERTTREPTTKSTTTTHETTABPAKKTHTHEP	-----	-----	1090
QY	711	TTPKGTA-----	-----PTTPKEPAPTTPKEPA-----	-----PTTPKGTAPTTPKEPA-----	749
Db	1091	TTQKSTTLRIEETPRTRKSSATAKTREPTTKRETERTTKEPTTRKKTTHKTTEEPTTKK	-----	-----	1150
QY	750	---PTTPK-----	-----KPAPKELAPTTPKGTSTTSDDKAPTTPKETAAPTTPKEPA-----	-----	794
Db	1151	TTHEPTTKKSTTLKPTSE-----	-----PTTRK-----	-----TSTTKTREPTTKRKTTERTTKEPTTRKKTTHKTTEEPTTKK	1205
QY	795	-----	-----PTTPK-----	-----KP-----	823
Db	1206	KTTEEPTTNKTTTKTTHETPTTKKSTTLKPTSEPTTRKSTTTKTREPTTKRKTTERTT	-----	-----	1265
QY	824	EP-----	-----TTIHKSPDE-----	-----SIPELSAEPTPKALENSKPEGVPTTKTPAAKPEMTT	873
Db	1266	EPTRKKTTHKTTEEPTTKKTTHETPTTKK-----	-----STTLKPTSEPTTRKSTTTKTTEEPTTK	-----	1324

QY	874	AKDKTTERDLR	-----	894
Db	1325	TKRVTTERTREPTTRKTTTHKTTTEPTTKTTTKTTTHKTTTEPTTKTTTLKLTBETPTTRKT	-----	1384
QY	895	---KMTKKTAT---	TTTKTTTSSKLTATTTQVTSITSTTQDTPFPFKITTTTKT-----	939
Db	1385	STTKTREPTTRKRVTTERTTREPTTRKT--	THKTTTEPTTKTTTKTTTHKTTTEPTTKKSTT	1442
QY	940	LAP	-----	977
Db	1443	LKPTTEPTTRKTSITTKTREPTTRKRVTTERTTREPTTRKTTTHKTTTEPTTKTTTKTT	-----	1502
QY	978	PKPQ	-----	991
Db	1503	HEPTTKSTTLKPTTEPTTRKTSITTKATREPTTRKRVTTERTTREPTTRKTTAHTKTEEPT	-----	1562
QY	992	-----	STKKPKMPRVKPKTTTPTRKTTSTMPELNPTSRIAEAMLOTT---	1035
Db	1563	TKTTHKTTHEPTTKSTTL	-----	1618
QY	1036	-----	TRPNQTPNSKLVEVNPKSDAGGAEETP-----	1072
Db	1619	TKRTTAEMSTTNOEPTSVETITNSSNQSNSTTESITTEOHVHHHHHHIHYE	-----	1675
QY	1073	FMPEVTPDMVLPVNPQGIINPMLSDETNI	CNGKPVGLTTLRNGTLVAFRGHYFWML	1132
Db	1676	LGPSILP	-----	1687
QY	1133	SPFSPSPARRIT	TEWVGIPSDITVTRCNCEGKTFPFKDSQYWRFTTNDIKDAGYKPKIP	1192
Db	1688	-PLPLPLP	-----	1707
QY	1193	KGFGGLGQIVAA	LSTAKYNWPESVYFFKRGGSIQQYIVKQEPVQKCPGRRPALNPVY	1252
Db	1708	-----	PPLPTA-----	1725
QY	1253	GEMTQVRRRPERAIG	SPQTHTIRIQVSPARLAYQKGVLHNEVKVSIILWRGLPNVVTSA	1312
Db	1726	-----	-----	1733
QY	1313	ISLPNIRKPD	1322	
Db	1734	ISLPEISLPN	1743	
RESULT 10				
AAS64673				
ID	AAS64673	PRELIMINARY; PRT; 1761 AA.		
AC	AAS64673;			
DT	01-APR-2004	(TrEMBLrel. 27, Created)		
DT	01-APR-2004	(TrEMBLrel. 27, Last sequence update)		
DE	01-APR-2004	(TrEMBLrel. 27, Last annotation update)		
DE	CG33300-PA	(Fragment)		
GN	CG33300.			
OS	Drosophila melanogaster	(Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
FP	SEQUENCE FROM N.A.			
EX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,			
RA	Abiril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,			
RA	Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova N., Borchen M.P., Bouché J., Boudreau N., Boudreau N.,			

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Meunier S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RA Science 287:2185-2195(2000).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=2426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RA melanogaster euchromatic genome sequence.";
RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bertencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RA systematic review.";
RA Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the *Drosophila melanogaster* euchromatin:
RA a genomics perspective.";
RA Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RN SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RA EMBL; AB003626; AAS64673.1; -.
RN NON_TER 1
RN SEQUENCE 1761 AA; 197521 MW; 62A5E16B6241B3F4 CRC64;
SQ

Query Match 18.6%; Score 1349.9; DB 2; Length 1761;
Best Local Similarity 25.2%; Pred. No. 3 3e-19;
Matches 472; Conservative 117; Mismatches 400; Indels 881; Gaps 91;
QY 47 AQCCKYDKCCPDYEFCAEVHNPSPSSKKAPPPSGASOTIKSTTKRSPKPNKKTKK 106
DB 161 AETKK-----STHNP-----TKKSTTLRTTEPTTSSST 191
QY 107 VIESEEITEHVSVE--NOESSSSSS-----SSSSSTIWKI-----KS 143
DB 192 AKTTREPTTKRETTERTTQEPSTSKTTHETTAEPATKKTTHETPTQKSTTLRTTEPTT 251
QY 144 SKNSANRELQKKLVKDNKNKTKKPTPKPPVDEAGSLGNDGDFKVT-----PDT 197
DB 252 RKGSTA-----KTTREPTK-----RETTERTTKEPST 279
QY 198 S-TQHNKVSPTKIITAKPINRPSLPNSDSKETSLSLVNKETTV-----243
DB 280 SKTTTHTHE-----TTAEPATKKTTHETPTQKSTTLRTTEPTT 329
QY 244 ---ETKETTT-----TNKOTSDGKEKTTSAKETQ-----SIEKTSADLAPTS 284
DB 330 TKRETTERTTQEPSTSKTTHETTAEPATKKTTHETPTQKSTTLRTTEPTT--SSTA 387
QY 285 KVLAKPTPKAET---TTKGPALATTPKEPTTTPKEPAS--TTPKEPT-----PT 328
DB 388 KTTREPTTKRETTERTTKEP--STSKTTHETAAEPATKKTTHETPTQKSTTLRTTEPTT 445
QY 329 TIKSA-----PTTPKE-----PA-----PTTKSA-----348
DB 446 TRKSSTARTTREPTTKRETTERTTQEPSTSKTTHETTAEPATKKTTHETPTQKSTTLRT 505
QY 349 ---PTTPKEPATTTTKEPATTPKEPATTTTKEP-----APTTKSA---PTTP 391
DB 506 TEEPTTRKSSAKTTRE--PTTKRETTERTTQEPSTSKTTHETTAETAKKTTHETPTQ 563
QY 392 KEPA-----PTTPKXPATTPKEPATTPKEPTTTPKEPA-----PTTKBPA---434
DB 564 KSTTLRTTEPTTRKSSAKTTRE--PTTKRETTERTTQEPSTSKTTHETTAEPATKKT 621
QY 435 ---PTTPKEPA-----PTAPKXPATTPKEPATTPKEPATTTKBSPT-----TPK 479
DB 622 THEPTQKSTTLRTTEPTTRKSSAKTTRE--PTTKRETTERTTQEPSTSKTTHETTA 679
QY 480 EPATTTKSAPTTKKEPA-----PTTKSA-----PTTPKXBSPTTKKEPA-----520
DB 680 EPATKKTTHETPTQKSTTLRTTEPTTRKSSAKTTREPTTKRETTERTTKEPTTRKTTT 739
QY 521 -----PTTPK-----EPA-----PTTPKXPATTPKEPA---544
DB 740 HKTTREPTTKTTHETPTKSTTLKPTTEPTTRKSTTKTTREPTTKRKTERTTKEPTT 799
QY 545 -----PTTPKEPATTTTKPA--PTAPK-----EPA-----PTT 571
DB 800 KKTTHKTTTEPTTKN-----TTTKTTHETPTTKSTTLKPTTEPTTKKSTTLKPTTEPTT 855
QY 572 PKETA-----PTTPKKT-----PTTPKKT-----PTT-PEKLAPT 595
DB 856 KRETTERTTQEPSTSKTTHETTAEPATKKTTHETPTQKSTTLRTTEPTTRKSSAKTT 915
QY 596 PEKPAPTTPEELAPTTPBEPT-----PTTPE-----EPA-----PTTP 628
DB 916 RE---PTTKRETTERTTKEPTTRKTTTHKTTTEPTTKTTHETPTTKSTTLKPTTEPTT 972
QY 629 KAAANTPKEPATTPKEPATTPKEPA-----PTTKETA-----PTTPKG 670
DB 973 KTSITTKTTR--PTTKRKTTEKTKEPTTRKTTTHKTTTEPTTKTTHETPTTKS 1030
QY 671 TA-----PTTLKEPATTPKXPAPKELA-----PTTKPEPTSTSDKPA-----P 710
DB 1031 TTLKTEPTTRKTTTKTTREPTTKRVTERTTREPTTSKTTTHETTAEPATKKTTHETP 1090


```

QY 711 TTPKCTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKEPA--- 749
Db 1091 TTQKSTLIRITEEPTTRKSSAKTIRTEPTTKREITERTKTEPTTKTHKTEPTTK 1150
QY 750 -----PTTPK-----KPAPKELAPTTKGTSTSDKPAPTTKETAPTTPKEPA----- 794
Db 1151 TTHTPTTKSTLKPTEE---PTTRK---TSTTKUTREPTTKRKTERTTKETPTTKTTH 1205
QY 795 -----PTTPK-----KP-----APTTPPTPTSEVSTPTTK 823
Db 1206 KTTREPTTKTKTTTHPTTKSTLKPTEPTTRKTSITTKTREPTTKRVTERTR 1265
QY 824 EP-----TTIHKSPDE-----STPELSAETPKALESPKPEPGVPTTKPAATPEMTT 873
Db 1266 EPTTKTTHKTEPTTKTKTTKTTHEPTTK-SITLKTEPTTKSTTKTTRPT 1324
QY 874 AKDKTERDLR-----TTPETTTAAP----- 894
Db 1325 TKRVTERTRPTTKTTHKTEPTTKTKTTKTTHEPTTKSTLKPTEPTTKT 1384
QY 895 ---KMTKETAT---TTEKTESKITATTQVTSITTDTPFKLTLTK-----TT 939
Db 1385 STTKTREPTRKVTTERTRPTTKT---THKITEPTTKTKTTKTTKTT 1442
QY 940 LAP-----KVTTTK-----KTIITHEIMMKP-----EETAKPKDRATNSKAT 977
Db 1443 LKTEPTTRKTSITKTTREPTTKRVTERTRPTTKTTHKTEPTTKTKTTKTT 1502
QY 978 PKQO-----KP-----TKAPKKPT----- 991
Db 1503 HEPTTKSTLKPTEPTTKTSTTKATREPTTKRVTERTRPTTKTTHKTEPT 1562
QY 992 -----STKKPKTMRVRKPTTPTPRKMTSTMPELNPTSRIAEAMLOTT--- 1035
Db 1563 TKKTTTKTTHPTTKSTL---KTEPTTRKTSITKTTREPTTRKTSVKTADQ 1618
QY 1036 -----TRPQNTNSKLVNPKSBDAGAGETP-----HMLLRPHV 1072
Db 1619 TTKRTTAEMSTTNOEPTSVTTTSSNQSTTTESITTEHQVHHHHHHIYH---KPAD 1675
QY 1073 FMEVTPDMVDYLRVNVQGIINPMSLDEINICNGKPEVDGLTLRNGTLVAFRGHYFWM 1132
Db 1676 LGSILP-----LPDLP----- 1687
QY 1133 SPSPSPSPARRITEVWGIPSIDVTRCNCEGKTFPFDQSYWRFTNDIKDAGYKPIF 1192
Db 1688 -PLPLPLP-----WP-PLPLPEI-----PLPL- 1707
QY 1193 KGGGLTGQIVAAALSTAKYKNWPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNV 1252
Db 1708 -----PPLPTA-----LPPLPLPLP----- 1725
QY 1253 GEMTQVRRRFRERAGPSQTHIRIOYSPARLAYQDKGVHLNEVKVSIILWRGLPNV 1312
Db 1726 ----- 1733
QY 1313 ISLPIRKED 1322
Db 1734 ISLPIELSLN 1743

```

RESULT 11

```

Q6DNC4
ID Q6DNC4 PRELIMINARY; PRT; 251 AA.
AC Q6DNC4
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Proteoglycan 4 (Fragment).
GN Name=PRG4;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match
Best Local Similarity 18.2%; Score 1324.3; DB 2; Length 3150;
Matches 460; Conservative 135; Mismatches 481; Indels 1487; Gaps 89;

QY 69 PTPSP-----SSKXAPPPSGASQIKSTTKRSPPPNKKTKKVIIEETEEHSV--- 119
Db 771 PTPSPSTTTWSSASTPSPS-----TPGTTRTPTRTSTESTDTTMSASTSEPTTPG 826

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jones A.R., Hughes C.E., Flannery C.R., Caterson B.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY653038; AAT74746.1; -.
FT NON_TER 251
FT NON_TER 251
SQ SEQUENCE 251 AA; 28659 MW; 60B1B474B62A743B CRC64;

```

```

Query Match
Best Local Similarity 18.3%; Score 1333; DB 2; Length 251;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1107 GKPVDGLTLRNGTLVAFRGHYFWMSPFSPSPARRITEVWGIPSPIDTVTRCNCEGK 1166
Db 1 GKPVDGLTLRNGTLVAFRGHYFWMSPFSPSPARRITEVWGIPSPIDTVTRCNCEGK 60
QY 1167 TFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGG 1226
Db 61 TFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGG 120
QY 1227 IQQYIYKQEPVQKCPGRRPALNVPGEMTQVRRRFRERAGPSQTHIRIOYSPARLAY 1286
Db 121 IQQYIYKQEPVQKCPGRRPALNVPGEMTQVRRRFRERAGPSQTHIRIOYSPARLAY 180
QY 1287 QDKGVHLNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYVAFSKQOYINIDVPSRTARA 1346
Db 181 QDKGVHLNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYVAFSKQOYINIDVPSRTARA 240
QY 1347 ITTRSGQTLTK 1357
Db 241 ITTRSGQTLTK 251

```

RESULT 12

```

Q7PMD5
ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANG00000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAA13969.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF_recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT NON_TER 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CE8C839 CRC64;

```

```

Query Match
Best Local Similarity 17.9%; Score 1324.3; DB 2; Length 3150;
Matches 460; Conservative 135; Mismatches 481; Indels 1487; Gaps 89;

QY 69 PTPSP-----SSKXAPPPSGASQIKSTTKRSPPPNKKTKKVIIEETEEHSV--- 119
Db 771 PTPSPSTTTWSSASTPSPS-----TPGTTRTPTRTSTESTDTTMSASTSEPTTPG 826

```

```
QY 120 -----GENQSSSSSSSS-----SSSIIWKI 141
Db 827 TTRTTPTSTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSAST---P 883
QY 142 KSKNSAANRELOKKLKVNDKKNRKKKPTPKPPVVDKAGS-----183
Db 884 EPTSTSGTTR-----TTPRPTDITMSSASTPEPSTTGGTTRTTPRPT 929
QY 184 ---GLDNGDFKVTTPDSTTQHN-----KVSTSKITAKP 216
Db 930 STSTDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTP 989
QY 217 INP-----RPSLPPNSDTSKETSUKNKETTVEIK 246
Db 990 TRPTSTESTDITMSSASTPEPSTKPGTTRTTPRPTSTESTDITMSSASTPEPSTTGGT 1049
QY 247 ETWTTNKQSTDGKETS-----KETQSEKTSKADLAP---282
Db 1050 RTTPT-RPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPST 1108
QY 283 ---TSKVLAKPTPKAET-----TTKGPALTTPKPTP-----311
Db 1109 TPGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTP 1168
QY 312 -----TTPKEPAS-----TTPKEP-----325
Db 1169 EPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSS 1228
QY 326 -----TPTTIKSAPTP-----KEPA--PTTIKSAPTP-----352
Db 1229 ASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITM 1288
QY 353 -----TTPKEA--PTTIKSAPTP-----KEPA-----TTPKAPAT 374
Db 1289 SASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTD 1348
QY 375 -----TTPKEA--PTTIKSAPTP-----KEPA-----TTPKAPAT 405
Db 1349 TTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTST 1408
QY 406 -----TTPKEAPTTPKEPTTTPKEAPT-----TKEPAP-----TTPKE 440
Db 1409 ESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPR 1468
QY 441 PAPT-----APKKPAP-----TTPKEPAP-----T 460
Db 1469 PTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRT 1528
QY 461 TPKAPATTPKEPSTTPKEPA-----PTTIKSAPTT-----TKEPA--PTTI 501
Db 1529 TPTPTPTTETMSSASTP-EPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGT 1587
QY 502 KSAPTTP-----KEPSTTPKEPAPTTPKEPAP-----529
Db 1598 RTTPTPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTT 1647
QY 530 -----TTPKAPAT-----TTPKEPAPT-----TTPKEPAPT 554
Db 1648 PGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASSSESTPEPS 1707
QY 555 TTKKPA-----PTAP-----KEPAPTTPKEAPTTPKLTPT-----586
Db 1708 TTPESSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDIT 1767
QY 587 -----TPE-----KLAPT-----TPE-----597
Db 1768 MSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTES 1827
QY 598 -----KPAPTTPKEAPTTPKEAPT-----TPE-----621
Db 1828 TDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPT 1887
```

```
QY 622 -----EPAPTTPKAAAPNTPKEPAP-----TTP 644
Db 1888 STESDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTP 1947
QY 645 KKPAP-----TTPKAPAPTTPKETA-----PTTKGAPTTP 675
Db 1948 TRPTPTETMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRT 2007
QY 676 LKEPAP-----TTPKAPKELAPT-----TKEPTST---703
Db 2008 PTRPTSTERDITMSSAYTPPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGG 2067
QY 704 -----TSDKPAPTTPKGTAPTTPKEP-----724
Db 2068 TRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGG 2127
QY 725 ---APTTP-----KEPAPTTPKGTAPTTPKEPAP-----750
Db 2128 TTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPST 2187
QY 751 -----TTPKKAPKELA-----762
Db 2188 GTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPE 2247
QY 763 PTTTKGPTSTSDKPAPT-----TPKE-----TAPTTPKEPAP-----795
Db 2248 PSITPGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPE 2307
QY 796 -----795
Db 2308 SASTPESTTPELLQOQHLPIPCRLVCQPLLRSHRHLVQPELLQOQHLPIPCROQ 2367
QY 796 -----TTPKKAPAT-----804
Db 2368 PLRSHRHLVQPELLQOQHLPIPCRLVCQPLLRSHRHLVQPELLQOQHLPIPCROQ 2427
QY 805 ---TPE-----TTPPTSEVSTTPPTKPTTIHKS---831
Db 2428 ASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTD 2487
QY 832 ---PDESTPELSAEP-----TPKALENSKPEP-----855
Db 2488 TMSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTD 2547
QY 856 -----GVPTTKTP-----AATKPEMTTT-----873
Db 2548 TTMSSASTPEPSTTGGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTDT 2607
QY 874 -----AKDITTRDURTP-----ETTAAPKMTKE 899
Db 2608 MSSSMSSASTPEPSTTGGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTES 2667
QY 900 TATT-----TEKTTESKITATTTQVTTTQDTPFKITTLKTTILAP---942
Db 2668 TDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPT 2727
QY 943 -----KVTTTKITITTEIMNKPEETAAPKDRAN-----972
Db 2728 TPTDSTMSSMSSASTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPT 2787
QY 973 -----SKATPKPO-----981
Db 2788 TPTDITMSSMSSASTPEPSTTGGTTRTTPRPTDITMSSMSSASTPEPSTTGGTTR 2847
QY 982 ---KPT-----KAPKPT-----STKKPKTMP---1000
Db 2848 TTPRPTDITMSSMSSSVTTPGTTRTTPRPTDITMSSMSSASTPEPSTTGGTTPGT 2907
QY 1001 KVRPKPTTTPRKTSTW-----PELNPT-----SRIAEAMLO-----1033
Db 2908 RTTTPRPTDITMSSMSSASTPEPSTTGGTTRTTPRPTDITMSSMSSASTPEPST 2967
QY 1034 -----TTRPNQTPNSKLVEVNPKEADAGAGETPHMLLRHVFMPEVTPDMDYLP 1085
```

```

Db 2968 TTPGTRTPTPTDSTMSASATPKPSSTPGTTRTP-----TP----- 3011
QY 1086 RVNQGIIINMLSDNETNLCKNPKVDGLTILRNGTLVAFRGHYFWMLSFFSPPARIT 1145
Db 3012 -----TDSMSSSSSACTPEPTTGTTRT-----TPTR----- 3041
QY 1146 EVMGIPSPIDVFTTRCNCCEGKTFEFKDSQYWRFTNDKAGYKPIFKGFGGLTQIYAA 1205
Db 3042 -----PTPTDIT-----MSSASTPEP----- 3057
QY 1206 LSTAKYKNWPSVFFKRGGSIQYIYKQBPVQKCGRRPALNVPGYGMTQVRRRREF 1265
Db 3058 -STT----- 3060
QY 1266 AIGSQTHIRIQXPAPLAYQDKGVHLNEVKVSLMRGLNVNVTSAISLNRKPDGYD 1325
Db 3061 ---PGTWT---TPT-----RPTDSTMSMSSSASTPE--- 3089
QY 1326 YYAFSKQYINIDVPSRT-----ARAITRSGQTLISKV-W 1359
Db 3090 -----PSTPGTTRTPTTPATISSRNPSSTSW 3118

RESULT 13
Q6SSE6
ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN Name=SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AASG0704.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1218; PSTLXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 17.0%; Score 1237.4; DB 2; Length 3409;
Best Local Similarity 14.0%; Pred. No. 2.5e-16;
Matches 396; Conservative 166; Mismatches 471; Indels 1796; Gaps 79;

QY 1 MAWKTLPI-----YLLLLLSV----- 16
Db 50 LAWVTPVSDALQFPVNVSVTSSAATDAFSAIYIGALLDLVNWVPWQNCINDTRRSYD 109
QY 17 ----- 16
Db 110 APWPSRCALPAVQYGYDEYDIDATVYSGNSLRPFYSTCRYPASSDYAFLEFYMPPWD 169
QY 17 -----FVIQVSSQ-----ELSCGR-----CFE 35
Db 170 DFGIYQPVLDGFMWVGFTVNTVDSNLKWDIPAWTAQGAWLGGQDPRDAVWFGTHYCSW 229
QY 36 SPERGECDDAQCKYD----- 53
Db 230 PFV---EC---SSCENYDIADPYDPDKIANGIVPAVITALDFRNASLIYYLDFGMPFAG 283
QY 54 -----KCCPD----- 58
Db 284 SLDNWWYLNAYNFIGGPIPANLPTLLPSLOHLALDHCRAITPDVRGTASLQYGWQOYPST 343

```

```

QY 59 ---YE----- 60
Db 344 GQPEYCSGDVGTSGDTEYVISGMPDEWGDVAAASDSSLPWANLRTVLSNQALYG 403
QY 61 -----SFAEVHN----- 68
Db 404 PIPEGLRSASSISSRWLRQGNTELCGLPEFAAPINSLLYLGTGLGTHTWVHDPADHTQ 463
QY 69 -----PTSPFS-----SKKAPPPSGASQI-----KST 91
Db 464 GGECLAPPPSPSPSPRPRPPPLPPSPPPPLLPSPVPVPPSPSPSPSPSPSPSPSP 523
QY 92 TKRSPKPNKKTKKVIESEBITSE-----H 117
Db 524 PPLPPSPSPPTPVARCICQVGICDPSPMPPSPPPPPPPPPPPPPPPPPPPPPPP 583
QY 118 SVSNQSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKKNRTKKTKTKPP 177
Db 584 PPSDPSPASS-----VPPSPEPP- 602
QY 178 VDEAGSLDNGDFKVTFTDTSTTOHNVKSTSPKITTAKPINRPSLPPNSDTSKETS 237
Db 603 -----SPKPPSPAPPSPAPPSP- 620
QY 238 NKETTIVTKETTTNKOTSTDGKEKTTSAKETOSIEKTSAKOLAPTSKVLAKTPKAB 297
Db 621 -----PPSPAPPSPAPPS- 633
QY 298 TKGPALTTPKEPTTTPKEPAST-----TKPEPTTTIKSAPTTTKEPAPTTTKA 350
Db 634 ---PAPPSPQPPFPVPPQPPSPVPPSPKPPSPAPPSPVPPSPAPPSPAPPSPAPP 690
QY 351 TPKEPAPTTTKBPAPTTK-----EPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 405
Db 691 SPAPPLPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 750
QY 406 -----TPKEPAPTTKEPTTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTT 448
Db 751 PSPEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 810
QY 449 PAPTTPKEPAPTTTPKEPAPT-----TTKPSPTTPK 479
Db 811 PEPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 870
QY 480 EPAPTTTKSAPTTTKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBPAP- 529
Db 871 PPSPEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 929
QY 530 -----TTPKBPAPTTTPKEPAPTTTPKEPAPTTTKKAPATAPK----- 565
Db 930 PPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 989
QY 566 -----EPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 610
Db 990 SDPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1048
QY 611 -----TPEPTTTPKEPAPTTTPKAAAPN-----TPKEPAPTTTPKEPAPTT 653
Db 1049 PPSPEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1108
QY 654 EPAPTTTPKEPAPTTTPKGTAPTTL-----K 677
Db 1109 SPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1168
QY 678 EPAPTTTPKBPAPKELAPTT-----TKEPTSTTSKDKAPATTP 713
Db 1169 PPAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1228
QY 714 KGTAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKBPAPKELAPTT-TTKGPT 772
Db 1229 APPSP-APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1283
QY 773 TSDKBPAPTTTPKEPAPTTTPKBPAPTTTPETPPPT-TSEVSTSTTTTKEPTTI 831

```

Db	1284	APPSPAPPSP---APPSPSPAPPQPPSPVPSPAPPSTTPPAPAPAAALPLPPSPA	1340
Qy	832	POESTPELSAETPKALE-----NSPKPFGV---PTTKIP	863
Db	1341	PLPVPSPAPSPPLRPQPTQTPAMPSPSPAPPSPAPPSPPPPTTPPLAP	1400
Qy	864	-----AATKPE-----MTTAA--	874
Db	1401	LPDCTLLAQAALLSIPDAANSVVFVVSAGLFTSVAPSTPPPELLASFCTVCSCOLTATAIS	1460
Qy	875	-----	874
Db	1461	LVGSSRGNNITNGSSGNGYNGGDAALQRIAPGNSTDRGSSGSSGSSWPGTETA	1520
Qy	875	-----KDKTERDL-----RTTP-----	887
Db	1521	AEWAVDAVQDGTQLQLSIGSVTYRTVVVDRTTPSVSGNVTLSANRIKQEPSAVGEASL	1580
Qy	888	-----ETTAAPKMT-----KETAT	902
Db	1581	NALGSKQAMLLTISFSEVPFAFDPAASLIIVTGALVAEWVAAADKMTFVYLAUTLPAELVA	1640
Qy	903	TTTKTTTESKIT-----	913
Db	1641	TAAGSSSSGTSRSGNGNGTAAAAAAAAPPAGTTGRRRALQOQAAAAAPPASGSSSL	1700
Qy	914	-----ATTQOV	919
Db	1701	SGRATANQOQHVHFLPATAYADAARNPGRNDLSLVELTQNAVASPAVGEALATTARV	1760
Qy	920	TSTTTQDTPFKITLTKTTLAPKVT-----	945
Db	1761	TAATYP-----AVAAATTLVAASSSFAQIRAKGSLLOGSYHIQMLTMSLYLASRGV	1813
Qy	946	-----TTKKTITTE-----INNK	959
Db	1814	GREYGEVAVFKYAVLGVKNLGPAAEMPTNEKEVTAABQARQVCGDLWPIGNDLLGGS	1873
Qy	960	PEETAKPKDRATNSKATTPKPKP-----	983
Db	1874	SNTTASGSSGSSSSNSP-PRRPPPPPAAGSTGLLFSNADASPPLAVATPALPAPLPS	1932
Qy	984	-----TYAPKKPTSTKKPTMRVRKPKTTPPKMTST--MPELN-----	1022
Db	1933	TIAAATAAP-----PRLPSP-PPPAVSGSTGVLPRHLQOMLQPPAAAV	1977
Qy	1023	-----PFSRAEAMLOTTTPNQTPNSKLYE-----	1048
Db	1978	AAPPPPPASSALVQLQSPPPPPPSQLLQOASATVSDMQDLLYTLVVAAMLITAVAA	2037
Qy	1049	-----VNPKSEDAGAGETPHMLLRPHVFMPEVTPMDYLRVFNQGIINP	1096
Db	2038	GLLIAAVLYRLVSP-----PHPELA-----	2064
Qy	1097	MLSDETNICGKPDGDTTLNGLTLVAFRGH-----	1127
Db	2065	-----TTIAGLILVALTYFCNALGGAADMHGSKTAAYCVLITAVVP	2107
Qy	1128	-----YFWMLSPFSPARRIFEVNGIP-----SPDITVFTRCN-----	1162
Db	2108	YAAFLWLWLA-----LARAMVQFTLVEPMTTSSVASSPSPAFERADTSRPSST	2157
Qy	1163	-----CEGKTPFFKDSQY-----WR-FT	1179
Db	2158	VSIATAADIGTATGGGGAASGACADVGAAMAAPATVAAGADVPSEDDRYARGPHWKQF-	2216
Qy	1180	NDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVYFFKGGSIQYIYKQFPVK	1239
Db	2217	-----DGVLPPTVASGVGGGVVPL-----PPLVA	2243
Qy	1240	CPGRRPALNYP-----VYGMTQVRRRRRFFERAIGFSQTH	1273

```

QY 59 ---YE----- 60
Db 344 QPVEYCSGDVGTSGTEYVISGMIPOEWGDAVAAASDSSLPWANLRTVLSNQALYG 403
QY 61 -----SFAEVHN----- 68
Db 404 PIPEGLRSASSISSWRLOQNTLGCPLPEFAAPINSLLYLGTGLGTHTWVHDPADHTQ 463
QY 69 -----PTSPPS-----SKAPPPSGASQTI-----KST 91
Db 464 GGECLAPPPSPSPSPRPRLPPPLPPSPVPPSPSPSPSPSPSPSPSPSPSPSPSP 523
QY 92 TKRSPKPNKKTKKVISEBELTE-----H 117
Db 524 PPLPPSPSPTPVARCIOVGICDPSMPSPSPRPQPPSPPPPPPPPPPPPPPPPPPP 583
QY 118 SVSNQESSSSSSSSSSSTTIWKIKSSKNANRELQKKLVKONKNTKKTKTPPVP 177
Db 584 PPSDPSPASS-----VPPSPSP----- 602
QY 178 VDEAGSLDNGDFKVTTPDTSTQHKNVSTSPKITTAKPINRPSLPPNSDTSKETSITV 237
Db 603 -----SPKPPSPAPSPSP----- 620
QY 238 NKETTVETKETTINKOTSTDGKEKTSKAKTQSIKTSKDLAPTSKVLAKTPKAEIT 297
Db 621 -----PPSPAPSPAPSP----- 633
QY 298 TKGPALTTKPEPTTTKPEKAST-----TPKEPTTTIKSAPTTKPEAPTTTKGAPT 350
Db 634 ---PAPSPQPPSPVPPSPVPPSPKPPSPAPSPVPPSPAPSPAPSPAPSPAPSP 690
QY 351 TKPEAPTTTKPEAPTTK-----EPAPTTTKPEAPTTKGAPTTKPEAPTTKPPAPT- 405
Db 691 SPAPPLPPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 750
QY 406 -----TPKEAPTTKPEPTTTKGPAPTTKPEAPTTKPEAPT-----APKK 448
Db 751 PSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 810
QY 449 PAPTTPKEAPTTKPEAPT-----TTKEPSPSTPK 479
Db 811 PEPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 870
QY 480 EPAPTTKGAPTTKPEAPTTTKSAPTTKPEPSPTTKPEAPTTKPEAP----- 529
Db 871 PPSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 929
QY 530 -----TPKPPAPTTPKEAPTTKPEAPTTTKKPAAPTAPK----- 565
Db 930 PPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 989
QY 566 -----EPAPTTKPEAPTTPKLTTPTEKLAAPTPE--KPAPTTEELAPT----- 610
Db 990 SPDPSAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1048
QY 611 -----TPEPTTPTEPAPTTPKAAAPN-----TPKEAPTTKPEAPTTPK----- 653
Db 1049 PPSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1108
QY 654 EPAPTTKPEAPTTPKGTAPTTIL-----K 677
Db 1109 SPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1168
QY 678 EPAPTTPKKPAKELAPTI-----TKPEPTSTSDKAPATTP 713
Db 1169 PPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1228
QY 714 KGTAPTTKPEAPTTPKGTAPTTILKEAPTTTPKPAKELAPTI--TKGPTST 772
Db 1229 APPSP-APSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1283
QY 773 TSDKPAPTPKETAPTTPKPEAPTTKPAAPTTPPETPPT-TSEVSTPTTTKPTTHKS 831

```

```

Db 1284 APPSPAPSP-----APSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1340
QY 832 PDESTPELSABPTPKALE-----NSPKPVG-----PTTKTP----- 863
Db 1341 PPLPVPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1400
QY 864 -----AATKPE-----MTTAA----- 874
Db 1401 LPPDCTLLAQAALLSIPDAANSFVWSAGLFTSVAPSTPPELLASFCVCSQLTATALS 1460
QY 875 ----- 874
Db 1461 LVGSSRRNNYNGSSGNYNGSDAAIORIPAGNSTDRGSSSGSSGSSWGPGTETA 1520
QY 875 -----KDKTBERDL-----RTP----- 887
Db 1521 AEMAVDAVDQCTYQLQLSIGGVYTRIVVVDRTTPSVGNVTLSANRIKQEPSAVGEASL 1580
QY 888 -----ETTTAAPKVT-----KETAT 902
Db 1581 NALGSKOAMLLITISFSEPVPAFDPAAASLIVTALVAEWAADKMTFVLAMTLPAELVA 1640
QY 903 TTEKTTESKIT----- 913
Db 1641 TAAGSSSGTSRSGNGNGNTAAAAAAPPAGTTGRRRALOQOAAAAPPSPSSSSSL 1700
QY 914 ----- 914
Db 1701 SGAATANQOQRHVHFLPATAYADAARNPGRNDLSLSVELTDNAVSPAUGEALATTARV 1760
QY 920 TSITTTQDTPPKITTTKTTTLAPKVT----- 945
Db 1761 TAAATYP-----AVAATTLVAASSSFAQAIKAGSLLGSSVHIQMLTWSLVSARGV 1813
QY 946 -----TTTKTITTE-----IMNK 959
Db 1814 GREYGEVAFKXAVIGVKNLGPAAAMPNEKEVTAABQARQVGDLWPIGNDLLGGS 1873
QY 960 PEETAKPKORATNSKATTPKPKP----- 983
Db 1874 SNTTSSSGSSSSSSNSP-PRPPPPPPAGSTGLLFSNADASPPPLAVATPALPAPS 1932
QY 984 -----TKAPKPTSTYKPKMTRVRKPKTTPTRKMTST--MPELN----- 1022
Db 1933 TIAAATAAP-----PRLPSP--PPAVGSSTGVLPRLHLMQMLQPPAAAV 1977
QY 1023 -----PTSRIABAMLOTTTRPNQTPNSKLVE----- 1048
Db 1978 AAPPPPPASSALVLQSPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2037
QY 1049 -----VNPKSEDAGGAEGETPHMLLRPHVFMPEVTPMDYLPVRPNQGIINP 1096
Db 2038 GRLLAAVLYRLLVSP-----PPHFLA-----PPRLE----- 2064
QY 1097 MUSDETNINCNGKPVQGLTTLRNGTLVAFRGH----- 1127
Db 2065 -----TTIAGLILVALTYFSCMALGGPAADWHGSRTAAYCVLITIAVVP 2107
QY 1128 ---YFWMLSPFPSPPARRITEWGLP-----SPIDTVFTRCN----- 1162
Db 2108 YAAFLWLLA-----LARAMVPOFTLVEPMTTSSYASPSAFERADTSRPREST 2157
QY 1163 -----CEGKTFFPKDSOY-----WR-FT 1179
Db 2158 VSIATAADIGGTATGGGGAASGACADYGAAMAPATVAAGADVPSEDDRYARGPHWKF- 2216
QY 1180 NDIKDAGYKPIFKGFGGLTGOIQAALSTAKYKNWPESVYFFKRGGS:QQYIYKPEYOK 1239
Db 2217 ---DGVLPPTVASGVGGGGVVP-----PPLVA 2243
QY 1240 CPGRPALNYP-----VYGMTQVRRRRRFAIGFSOTH 1273

```

Db 2244 LFGSGRHLPPLPPLPAAGAAATAAGGAATAPGSPROAAAGE-----DTH 2287

QY 1274 TIRIOVSPARLAQDKGVILHNEVKVLSILWR----- 1303

Db 2289 ---QFGP-----H-----WKFSVPDDGAATAGVGYDAAAGGAGGGGGRG 2324

QY 1304 ----- 1303

Db 2325 SRGSGTDGVRFGANGTRTPSDGAKGRSSHHGGGNGGNAMSSGAASFGASAGGGGAAAAASR 2384

QY 1304 ---GLP---NVVTSALSL----- 1315

Db 2385 LGSRSRSPGDDYNAMSSGAAMRGGGAGASRFGPAGSRAGSPPEGAGGAGVAGGANAMFAGT 2444

QY 1316 -----DAPPPLPAPGTVPPLPFANLLTSGADVRSRRALNABAVAAAAGGG 2504

Db 2445 AGRAGALSPADGSSRRPGSGSNEGSRQMSYGSNNAMTAGAVGVGGVRRPAGSLRNASD 2504

QY 1323 GVDYAFSKQDYNNIDVPS-----RTARAI----- 1347

Db 2505 G-----DAPPPLPAPGTVPPLPFANLLTSGADVRSRRALNABAVAAAAGGG 2550

QY 1348 ---TTRSG 1352

Db 2551 GGGGTSRSG 2559

RESULT 15

SLP1_CLOTH STANDARD; PRT; 1664 AA.

AC Q06852;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).

DE Name=olpB;

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 10682;

RX MEDLINE=9320931; PubMed=8458832;

PA Fujino T., Beguin P., Aubert J.-P.;

RT "Organization of a clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface.";

RL J. Bacteriol. 175:1891-1899(1993).

CC -!- SUBUNIT: Assembled into mono-layered crystalline arrays.

CC -!- SUBCELLULAR LOCATION: Cell wall.

CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X67506; CAA47841.1; -

DR PIR; T18262; T18262.

DR InterPro; IPR008965; Cellul_bind.

DR InterPro; IPR001119; SLH.

DR Pfam; PF00395; SLH; 3.

DR PROSITE; PS01072; SLH DOMAIN; 2.

DR Cell wall; Repeat; S-layer; Signal.

DR SIGNAL 1 28 Potential.

FT CHAIN 29 1664 Cell surface glycoprotein 1.

FT DOMAIN 36 763 4 X 156 AA approximate repeats.

FT REPEAT 36 191 1.

FT REPEAT 207 363 2.

FT REPEAT 409 565 3.

FT REPEAT 607 763 4.

FT DOMAIN 771 1377 approximate tandem repeats of T-P-S-D-E-P.

FT DOMAIN 1378 1449 Gly/Pro/Ser/Thr-rich.

FT DOMAIN 1453 1494 SLH 1 (incomplete).

FT DOMAIN 1495 1565 SLH 2.

FT DOMAIN 1566 1625 SLH 3.

FT DOMAIN 1626 1646 SLH 4 (incomplete).

SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match 16.4%; Score 1190.4; DB 1; Length 1664;

Best Local Similarity 20.8%; Pred. No. 5.9e-16;

Matches 408; Conservative 145; Mismatches 465; Indels 946; Gaps 82;

QY 4 KTLPIYLLLLLVFV-----LQOVSSQE 26

Db 6 KVLISILLTLLIITTSVNMSPFAEATPSIEMVLDKTEHVGDVITATIKVNNIRKLQAGY 65

QY 27 LSCCKGRCFESFERECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSK---KAPPPS 82

Db 66 LNIK-----FDPEVLQVDPATGEEFTDKSMP-- 92

QY 83 GASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENQESSSSSSSSSSSTIWKIK 142

Db 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGYNNLTAYK 131

QY 143 SS----- 144

Db 132 SSGIDHTGTIGEIFKVLKQNTSIRFEDTSLMPGAISGTSIFDMDAETITGVEVIQPD 191

QY 145 -----KNSAANRELQK-KLKVKD-----RTKKKT-- 172

Db 192 LIVVEAPLKDASVALELDTKVKVGDIITATIKIENKKNFAGYQLNKKYDPTMLEAIEL 251

QY 173 -----PKPPVVDVAGSGLDN----- 187

Db 252 ETGSAIAKRTWPVTGGTVLQSDNYGKTTAVANDVGAGIINFADYNSNLTKYRGTGVAET 311

QY 188 -----GDFKVTPTD-----STQHNKV----- 205

Db 312 GIIGKIGFRVLKAGSTAIRFEDTTAMPGAIEGTWFDWYGENIKGYSVVQGEIVAESEE 371

QY 206 -----STSPKITTAKPINRPSLP-----PNS-----DTSK----- 231

Db 372 PGEEPTPEPVTETPVDPTFTVTEEPVPSLPDYSVIMELDKTKVKVGDIITATIKIENM 431

QY 232 -----ETSLTVNKETTETKTTTNTKQTSTDGKEKTS----- 265

Db 432 KNFAGYQLNKKYDPTMLEAIELETGSAIAKRTWPTGGTV-----LQSDNYGKTTAVAND 486

QY 266 -----AKETQSIEKTSKOLAPTSKVLAKPTPKAETTT----- 298

Db 487 VGAGIINFADYNSNLTKYRGTGVAETGIIGKIGFRVLKAGSTAI-----RPEDTAMPG 541

QY 299 -----KGPALATP-----KEPTPT-TPKEPASTTPKEPTPT-- 328

Db 542 AIEGTWFDWYGENIKGYSVVQGEIVAEGEETPEPVTETPVDPTFTVTEEPVPSLP 601

QY 329 ----- 328

Db 602 DSYVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDKPKVLEAFNIETGDPIDEGT 661

QY 329 ----- 329

Db 662 WPAVGGTILKNRDYLPVTGAINNVKSGIINFADYNSNLTKYRGTGVAETGIIGNIGFRV 721

QY 330 IKSAPTTPK-----EPATTTKSAPTTPKEPAPT 359

Db 722 LKAEDTTIRFELESMPGSDIGTYMDVNLNRSIGYVVIQAPAKAAS-----DEPIPTD 776

QY 360 TKEPAPTTPKEPA-----PTTKEPAPTTPKSAPTTPKEPAPTTPKPAPTTPKEPAPT-- 413

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 27.7629 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-F
Perfect score: 7276
Sequence: 1 MAWKLPYLLLLSVFVIQ.....ARAITRSQGTLQKVVNCP 1363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1348.5	18.5	3020	2 A43932	mucin 2 precursor,
2	1190.4	16.4	1664	2 T18262	S-layer protein
3	1065.3	14.6	2187	2 T30826	nascent polypeptid
4	1061.4	14.6	7962	2 I38346	elastic titin - hu
5	1052	14.5	3570	2 T45025	mucin MUC5B, trach
6	1023.2	14.1	1274	2 T16251	hypothetical prote
7	1020.2	14.0	1489	2 T31108	cyst germination s
8	983	13.5	1367	1 S48478	glucan 1,4-alpha-g
9	968.5	13.3	6642	2 T29757	protein UNC-89 - C
10	965.1	13.3	3507	2 T34513	hypothetical prote
11	919.5	12.6	5762	2 A4819	proline-rich pepti
12	898.5	12.3	1188	2 S49915	extensin-like prot
13	897.8	12.3	2897	2 B48666	cell proliferation
14	890.1	12.2	1229	2 T25697	hypothetical prote
15	885.9	12.2	1344	1 A35175	mucin 1 precursor,
16	885.5	12.2	3256	2 A48666	cell proliferation
17	879.3	12.1	4135	2 T42629	tenascin-X - bovin
18	861.3	11.8	1151	2 T18535	high molecular mas
19	851.5	11.7	3942	2 T42730	Bassoon protein -
20	849	11.7	2142	2 T35098	MHC class III hist
21	846.1	11.6	2232	2 T34434	hypothetical prote
22	838.3	11.5	5262	2 T03454	ALR protein - huma
23	835	11.5	1832	2 T31113	mucin-like glycopr
24	831.3	11.4	3938	2 T42761	Bassoon protein -
25	826.5	11.4	1872	2 S36152	MHC class III hist
26	821.1	11.3	4548	1 S00657	apoptosis(a) (EC
27	817	11.2	4957	2 T03455	ALR protein - huma
28	815.8	11.2	5105	2 T32650	hypothetical prote
29	810.7	11.1	2774	2 A43359	microtubule-associ

ALIGNMENTS

RESULT 1

A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 09-Jul-2004
C:Accession: A49963; A45106; B43932; B33532; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A49963; MUID:94132002; PMID:8300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upst:
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:gi86395; PIDN:AAA59163.1; PID:gi86396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:gi86397; PIDN:AAA59164.1; PID:gi86398
A:Experimental source: colon
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.N.
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor:
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:gi88863; PIDN:AAA59875.1; PID:gi88864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A:Reference number: A33532; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:gi88873; PIDN:AAA36334.1; PID:gi88874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

30 810.3 11.1 1870 2 S37671 MHC class III hist
31 802.5 11.0 3381 2 T42389 vesican precursor
32 797.4 11.0 1611 2 T38236 hypothetical prote
33 791.4 10.9 2225 2 T26063 hypothetical prote
34 784 10.8 990 2 I51618 nucleolar phosphop
35 781 10.7 1630 2 A53577 ascites sialoglyco
36 778.7 10.7 4006 2 T09070 probable tenascin
37 774.7 10.6 5170 2 T15348 hypothetical prote
38 773.5 10.6 3968 2 A44265 trithorax homolog
39 772.2 10.6 4667 2 T20774 hypothetical prote
40 771.6 10.6 4549 2 T20771 hypothetical prote
41 771.1 10.6 1777 2 T34369 hypothetical prote
42 765.9 10.5 3869 2 A48205 All-1 protein +GTE
43 765 10.5 761 2 C84672 hypothetical prote
44 755.3 10.4 2649 2 T51023 hypothetical prote
45 753.7 10.4 5327 2 T13564 microtubule-associ


```

Db 2257 DDDR-----QENETWLCDFM-ATCKYNNNTVEIVKVECEPPPPMPTCSNGLQ 2302
QY 1052 ---KSEDAGGA-----EGETPHMLLRPHFMPVPTDM 1081
Db 2303 PVRVEDPGCWHWCDCYCTGWDGPHVYTFDGLYSYQGNCTYVLVE-----EISPSV 2356
QY 1082 DYLRVPNOGIIINPMLSDETNICNGKPVDTGLTLRN-----1118
Db 2357 D-----NFGVVIDNYHCD-----PNDKVSCTRLLIVRHETQEVLLIKTVHMMPMQVQ 2402
QY 1119 -----GTLVAFRG-----HYF-----1129
Db 2403 VQVNRQAVLPYKXYGLEVYOSGINVYVDIIPELGVLVSYNGLSFSVRLPYHRFGNNTKGQ 2462
QY 1130 -----WMLSPESP-----1138
Db 2463 CGTCNTTSDCILPSGEIVSNCERAAQWLVDPSKPHCHSHSSTTKRPVTVPGGGKT 2522
QY 1139 SPARRITEVMGIPSI-----DTVTRCN-----CEGKTFPFKDSQY-----1175
Db 2523 TPHKDCCT-----PSPLCQLIKDSLFAQCHALVPPQHYVDACVFDSCFMPGSSLECASLQA 2577
QY 1176 -----WR-----1177
Db 2578 YAALCAQONICLDNRNHTHGACLVCEPSHREYQACGPAEEPTCKSSSQNNNTVLVEGCF 2637
QY 1178 -----FTNDIKDAGY--PKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFPK--1222
Db 2638 CPEGTMYVAGFDVCVKTGCGVDNVPREG-----EHFEFD 2675
QY 1223 -----RGGS-----IQYIYKQEPVQKC-----1240
Db 2676 CKNCVLEGSGIICQPKRCSQKPVTHCVDGTYLATEVNPADCCNITVCKNCTSLCKE 2735
QY 1241 -----PGR-----RPALNYPVYGMTQ-----1257
Db 2736 KPSVCPPLGFEVKSQWVPGRCPPFYWCBSKGVCHGNAEYQP--GSPVYSSKQCCVCTDK 2793
QY 1258 -----VRRRRFERALGPSQTH 1273
Db 2794 VDNNTLLNVIATCHVPCNTSCSPGFELMEAPGECCKCEQTHCIIR-----PDNQH 2845
QY 1274 TIRIQSPARLAYODKG-----VLHNEVKYSILWRGLNPNVTSAISLPN-----IRKP 1321
Db 2846 VI---LKPQDFKSDPKNCTFFSCVKIHNQLISSV-----SNITCFNDFASICIP 2892
QY 1322 DGYDYVAFSKQYNYNDVPSRTARATTR-----S 1351
Db 2893 GSITF-----MPNGCCKTCFPRNETRVPVPCSTVPVTVTSYAGCTKTVLMNHC 2940
QY 1352 GQTLKSVWY-----NCP 1363
Db 2941 GSCGTFVWYSAKAQAALDHSCCKEETKSQREVVLSCP 2978

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; PMID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUC>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841

Query Match
16.4%; Score 1190.4; DB 2; Length 1664;

```

```

Best Local Similarity 20.8%; Pred. No. 2e-17;
Matches 408; Conservative 145; Mismatches 465; Indels 946; Gaps 82;

QY 4 KTLPIYLLLLLSVFV-----IQQVSSOE 26
Db 6 KVLISLLTLLLLISITSVNMSFAEATPSIEMVLDKTEVHVGVDTTATIKVNNIRKLAGYQ 65
QY 27 LSCKGRCPESPERGECDCQACKKYDKCCPDYEFCAEVHNPTSPSSK---KAPPS 82
Db 66 LNIK-----FDEVLQPDVPAIGSEFTDKSMP-- 92
QY 83 GASQTIKSTTKSPKPNKKTKKVIESEBITEHSHSVSENOESSSSSSSSSTTIWKIK 142
Db 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGYNLTAYK 131
QY 143 SS-----PKPPVVDGAGSLDN-----144
Db 132 SSGIDEHTIIGEIFKVLKKQNTSIRFDTLSMPGALSGLDFMDAETITGYEVIQPD 191
QY 145 -----KNSAANRELQK-KLKVKD-----NKN-----RTKKKPT-- 172
Db 192 LIVVEAPLKDASVALELDKTKVKGDIITATIKIENKMFAGYQLNKKYDPTMLEALEL 251
QY 173 -----PKPPVVDGAGSLDN-----187
Db 252 ETGSAIAKRTWPVTGTVLQSDNYGKTTAVANDVGAGIINFAPAEAYSNLTKYRETVABET 311
QY 188 -----GDFKVTTPDT-----STQHNK-----205
Db 312 GLIGKIGFVLKAGSTAIRFEDTTPMGAIECTYMFWDYGENIKGYSVVQGEIVAESEE 371
QY 206 -----STSPKITTAKPINRPSLP-----PNS-----DTSK-----231
Db 372 PGEETPEEPVTPVDPPTVTVEEPVSELPDSYVIMELDKTKVKVGDIIITAIKINM 431
QY 232 -----ETSLTVNKETIVETKETTITNTKQSTDKKETS-----265
Db 432 KNFAGYQLNKYDPTMLEALETGSIAKRTWPVTGTV-----LQSDNYGKTTAVAND 486
QY 266 -----AKETQSIEKTSADLAFTSKVLAKPTKAEITT-----298
Db 487 VGAGIINFABAYSNLTKYRETVABETGLIGKIGFVLKAGSTAI-----RFEDTTAMPG 541
QY 299 -----KGPALTP-----KEPTPT-TPKEPASTTPKEPTPT-- 328
Db 542 AIBGTYMFDWYGENIKGYSVVQGEIVAEGETPEEPVPTPTVPTPTTTEPVPSELP 601
QY 329 -----328
Db 602 DSYVIMELDKTKVKGSDVIIITIRVNNIKNLAGYQIGIKYDPKVLAFNIETGDPIDEGT 661
QY 329 -----T 329
Db 662 WPAVGOTILKNRDYLTGTVALNNVSKGILNFAAYVYVFDYREEGKSEDTGIIGNGFRV 721
QY 330 IKSAPTPK-----EPAPTTTKSAPTTKEPAPT 359
Db 722 LKAEDTTIRFEELSMPSGIDGTYMLDYNLRISGVYVWIQAPAKAAS-----DEIPTD 776
QY 360 TKEPAPTTKEPA---PTTKPAPTTTKSAPTTKEPAPTTPKPAPTTPKKEPAPT-- 413
Db 777 TFSDEPTPSDEPTPSDEPTPSDEPTSETPEEPITDTPSDEPTPSDEPTPSDE 836
QY 414 -----TPK-BPTPT-TPKEPAPT-TKEPAPTTKEPAPTAPKKA---PTTKPAPT- 460
Db 837 PTFSDEPTPSDEPTSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSE 896
QY 461 TPKEPAPTTKEPSPPTTKPAPTTTKSAPTTKEPAPTTTKSAPTTTKSAPTTTKKEPA 520
Db 897 TPEEPIPTDTPSDEPTPSDEPTPS---DEPTPSDEPTPSDEPTPSDEPTPEEPIPTDPSDE 953
QY 521 PTTKEPAPTTKKA-----PTTKPAPT-TPKEPAPTTTKKAPATPAK 565

```



```
QY 337 PKK-PAPTTTKSAPTPK-----EPAPTTTKEPAPTTTKEPAPTTTKEPAP----- 381
Db 980 PKKSPKPAASKTATPSPGVTAVPLEIPPCSKKAPKTAAPKESATSSSRAPKTAVS 1039
QY 382 -----TTTKSAPTTKEPAPTTPKK-PAPTTTPKE----- 409
Db 1040 KEIPSGVTAVPLEISLPLKETSKA--TPGKSASSPKRSPKTAGPKETPGGVTAVPP 1097
QY 410 ----PAPTTTPKEPT-----TTTPKE-----PAPTTTKE 432
Db 1098 EISLPPKETPQWATPNESLAASSQKRSKPTSPVKETPPGGVTAMPLEIPSAPOKAPKTAV 1157
QY 433 P-----AFTTPKEPAPTAPKAPATT----- 454
Db 1158 PKQIPTPBDVAITLAGSLSPKASKATAAPKEAPATPSGVTAIVSGEISPSPKKTSKTA 1217
QY 455 -----KEPAPTTTKEPAPTTTKE-----PSPTTTPKE-----PAPTTTTSAP 490
Db 1218 PKENSATLPPKRSKPTAAAPKE--TPTATSEGVTAIVSEISPSPTPASKGVPTLTTPKGAP 1276
QY 491 TTTKEPAPTTTTSAPTT--PKEPS--PTTTKEPAPTTTKEPAPTTTKEPAPTTT 539
Db 1277 NALAE-SPASPKVPKTAAPETSTTSQKIPKVGAPKEASATPPSKTPTKTAVPKETS 1335
QY 540 -PKEPAPTTTKEPAPTTTKEPAPTPAKE-PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTPE 597
Db 1336 APSEGVTAIVPLEIPSPKAPKTAAPKETPAPS--PEGATTAPVQI-PPSPKSKKAGS 1392
QY 598 KPAPTTPEELAPTTPEPTTTPREP--APTTPKAAP-----NTP 636
Db 1393 KETPTTP-----SPEGVTAAPLEIPISSKTSKVASPKETLVTTPSSKSLSQTVGPKETS 1446
QY 637 KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 694
Db 1447 LEGATAVPLEIPSHKAPKTVDPKQVPLTPSPKDAFTTLAE-SPSSPKK-APKTAAPPS 1504
QY 695 --TTTKEPTTSDKAPATTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 743
Db 1505 ERYTTTTPP-----EKPA-TPQKASGTTASKVPVPAETQEVAVSSRETPTVPAPVPKNS 1558
QY 744 TLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 803
Db 1559 SHAKTSKTIELKAPATLPSPTKSPKIPSSKK-----APRTSAP-----KEFP 1602
QY 804 TTPETPTTSEVSTPTTKEPTTIHKSPPDESPE--LSAEPKPALENSPKPGVPTTK 861
Db 1603 ASPSIKPVITSLAQT-----APSLQKAPSTTIPKENLAA--PAVLVSSKSPAAP-AR 1653
QY 862 TPAATKPEMTTAKOKTTERDLRTP-----ETTTA-----APKMTKETATTT 904
Db 1654 ASASLSP-----ATAAPQAPKEATTIPSCKAAATETPIETSTAPSLGAPKETSETSV-- 1708
QY 905 EKTESKITATTTQVSTTTQDTPPKITL-----KTTTLAPKVTTKTITTTTEIMNKP 960
Db 1709 -----SKVMSGPPKKASSKSKASTLPAATLPLSLKEASVLSPTATSSGK----- 1752
QY 961 BETAKPKDRATNSKATTPKQPKTPAKPK----- 989
Db 1753 DSHISPVSDACSTGTTT--PQASEKLPSKKGPTAFTEMLAAPAPESALAITAPIQSPGA 1810
QY 990 -----PTSTTKPKTMWR--VRKPKTPTTPKMTSTMTPELNPPTSRIAEAMLQTT 1035
Db 1811 NGSASSPKCPDPSSKKDKTGLPSAVALAPQTVPEK-----DTSKAETLLVSP 1860
QY 1036 TRPNQ----- 1047
Db 1861 AGSDCLHSPKGVSGQVATPIAAFTSDKVPPEAVSASVAPKAPAPASLTLAPSPVAPLP 1920
QY 1048 EYNPKSADAGGAEGETPHMLLRPHVPMBVTPDMDYLPRVPNGIILNPMLSDEINICNG 1107
Db 1921 PKQPLLESAPGVSLESPSKL-----PVPABEDELPL-----PLIPPEA-VSGG 1961
QY 1108 KPVGDGLTLRLNGLTVAFRGHYFWMILSPFSF-----PSPARRITEVMGIPSIDTFTFR 1160
```

```
Db 1962 E-----PQPIILNNPAPKPA-----GTPAPAPGA----- 1986
QY 1161 CNCEGKTTFFKDSQYWRFTNDIKDAGYKPKPIFKGG----- 1197
Db 1987 -----KQPVILKNKSGTSDSDSDESVPELEEQDSTOTA 2019
QY 1198 -LTGQIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVKQKCGRRPALNYPYVGEMT 1256
Db 2020 TQQAQAAAAE-----IDEEPVSKAQSR-----S 2044
QY 1257 QVRRRRFERAIGPSQTH-----TIRIQVSPARLAYQDKGVILHNEVKVSIILWRGLENNVTS 1311
Db 2045 EKKARKAMSKIGLQVGTGVTIR-----KSKNILF-----VIT- 2079
QY 1312 AISLNRKPKDGY-----DYAFSKQYQYINIDVPSTRARAITTR-SQITLSKVVYN 1361
Db 2080 -----KPDVYKSPASDTIIVFGEAKIEDLSQAQAAAAEKPKVQGEAVSNIQEN 2128

RESULT 4
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labett, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g10
C:Genetics:
A:Gene: GDB:TIN
A:Cross-references: GDB:I127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 14.6%; Score 1061.4; DB 2; Length 7962;
Best Local Similarity 15.3%; Pred. No. 3.3e-13;
Matches 429; Conservative 152; Mismatches 508; Indels 1716; Gaps 106;

QY 2 AWK-----TLPIYLLLLSVFVIOQVSSQ----- 25
Db 5411 AMERHLQDVTLKEGQCTMTVQ-----FSPVNVKSEFRNGRILKPKQGRHKTEVHKV 5463
QY 26 -----ELSCGRCFESPGRGECDCQCKKYDK-----CCPDYESFCAE- 65
Db 5464 HKLTIADVRAEDQGGYTCK--YEDLETSAELRIEAEPQFTKRIQNIIVVSEHQSATFEC 5520
QY 66 -----VHN----- 68
Db 5521 EVSFDDAIVTYKGPTELTSOKYNFRNDGRCHYMTIHNVTDPDDEGVSVIARLEPRGEA 5580
QY 69 -----PTSP-----PSSK 76
Db 5581 RSTAEIYLTKEIKLELPPDIPDSRVPIPTWPIRAVPEEIPPVVAPVPLLPTPEEK 5640
QY 77 KAPP-----PSGASQTI----- 88
Db 5641 KPPPRIEVTKAVKADKAVKPKEMTPREEIVKKPPPTTLIPAKAPEIIDVSSKAE 5700
QY 89 ----- 88
Db 5701 EVKIMTIITRKKEVQKEAVVEKQAVHKEKRVFIESFEEPYDELEVPYTFEPPEQPYE 5760
QY 89 ----- 88
Db 5761 EPDEYEEIKVEAKEVHEWEEDPEEGQYEREEGYDEGEWEAEYQREVIOVQKE 5820
QY 89 -----KSTTKRSPKPPN--KK-----KTKKVISEEEI----- 113
```

Db	5821	VVEESHKRVPAKVPKAPPPKVIKXVIEKIEKTSRMBEKQVQVTKVPEVSKVI	5880	Db	6873	PPAKVPVPKVEKRIILPKBEEVLPVEVTBEEPERISEBBIPEPPSIEEVEVAPP	6932
Qy	114	-----TEHSVSENQESSSSSSSSSSSTIWKI-----	141	Qy	552	-APTTTKKAPAPAP-----KEPAPTTPKET-----	575
Db	5881	QKPSRTPVQEEVIEVQVPAVHTKKWVISEKMPFASHTEEVSVTVPEVQKEIVTEEKIH	5940	Db	6933	RVPEVIKKAAPVAPPTVPKKVAPPAKVKSKKIPKVKVPVQKKEAPPAKVPVPPKKVPE	6992
Qy	142	-----KSSKNSAANRELQKLVKDNKNKNTKKP-----	171	Qy	576	-----	575
Db	5941	VAVSKRVPPKVPPELPEKAPBEVAVPIPKKVPAPPAKVPVPPKVPPEEKVPVPVK	6000	Db	6993	KKVLVPKKEAVPPAKGRTVLEEKVSVAFRQVVVVKERLELVVEAEVEEIEPEEPEHE	7052
Qy	172	-----TPKPPVDEAGSLDNGDKVTTPTDTSIQ-----HNKVSSTPKI	211	Qy	576	-----	575
Db	6001	KEPAAPPKVPVPPKVPPE-----KIPVPVAKKKEAPPAKVPVQKGVVTEEKI	6050	Db	7053	EYFEEGEFHEVEBFIKLEQHRVSEEHVHRVHVEVFEABEVEVEFKKAPKGPGEISE	7112
Qy	212	TTA-----KPINPR-----PSL-----	223	Qy	576	-----APTPKKLT-----PT-----TPE	589
Db	6051	TIVTOREESPPAVPEIPKKVPPEEKVPV-PRKEEEVPPPKVPALPKKVPPEEKVAVPV	6109	Db	7113	KIIPPKKPTKVVRKBPAPKVPPEVKIVVEKVRVPEEPVPFPKVPPEVLPKPEVVPE	7172
Qy	224	PPNSDTSKET-----	233	Qy	590	KLAPT-----TPEKPAPTTPEELAPTTPEE-----PTPTTPEEP-----APTTPKAAAPN	634
Db	6110	PVAKAPPPRAVSVKKTWBEKRFVABEKLFAVQVRVETRHEVSAEBSYSSEBEGV	6169	Db	7173	KKVPVPPAKKPEAPPKVPPEAPKVVPEKKVPVPPKPPVPTKVPEVPEKAAVPEKKVP	7232
Qy	234	SLTV-NKETTVETKETTNNKQSTDCKEKTSAK-----ETQSIEXT-----	275	Qy	635	-----TPKRPAPTTPKPE-----APTTPKPE-----APTTPKETATTTKGTAPTTLKE	678
Db	6170	SISVIREEEREAEVTEYVMEEPPEEVVVEEKLHIISKRVAEAEVTERQEKKIVL	6229	Db	7233	EALPPKPPESPPEVFBEPPEESPAP-PPKPEVPPVRVPEVPEVPEKKVPAP	7285
Qy	276	-----SAKDIAPTSKVLAKPPPKAETTTKGPALTTPKPTTTPKPE	317	Qy	679	PAPTTPKPE-----APKELAPTTTKEPTSTTSKPAPTTTPKGTAPTTPKPE-----	724
Db	6230	KPKIPAKIEBPPPAKVPPEAKKIVPEKKVPA-PVPKKE-----KVPPKVPPEEP	6277	Db	7286	-----PKKPEVTVKVPPEAKPEVVP-----EKKVPVP-----PKKPEVPT	7322
Qy	318	ASTTP-----KEPTP-----TTKSAPTTKBPATTTKS	347	Qy	725	-APTPPK-----EPAPTTPKGTAPTTLKPE-----AP-----T	751
Db	6278	KKVPPEKKVPKVKIMEEPLPAKVTEKHMQITOBEKVLVAVTKKEAP-PRARVPEPKR	6335	Db	7323	KVPEVPAVPEKKVPEALPPKPPESPPEVFBEPPEEVALBEPPEAEVVEEPEAPPOVTV	7382
Qy	348	A-----PTTPKBPATTT-----KEPAP-----	365	Qy	752	TPKKAPKELAPTTTKGTSTTSKPE-----APTTPKETAP-----TTPKPE-----	793
Db	6336	AVPEEKVLKPKKEEPPPAKVPFRKRVKVEKSVSEAPKREPOIKVTIMEEKERAY	6395	Db	7383	PPKNVPPEKKAP-----AVAKKBPPLPVKVPPEVPEVPEKKVLLVPPKPEADPAK	7435
Qy	366	-----TTPKBPATTT-----	369	Qy	794	APTTPKPE-----PAPTTPETPPPTTSEVSTPTTTKE-----PTTIHSPD	833
Db	6396	TLEZEAHSVQEEYEEYDYKEFBEYEPTEEDVQVEEYEEYEREBEHEVITEPE	6455	Db	7436	VPEVPEVPEKKVAVPKPEVPPAKVPPEVPPKVELEKPAVVPPEAESEPPPEVEPEPE	7495
Qy	370	EPAPT--TTKEPAPTTPKSAP-----TTPKPE-----PAPTTPKBPATTTKPE-----	410	Qy	834	ESTPE-----LSABTPKALENSPKBFG-----VPTTKTPAATKPEMTTAK-----	875
Db	6456	KPIPVKVPPEVPETKAPDAKVLKKAAPVEKVPVPIPKKLPKPPKVPPEEPKKVPEEK	6515	Db	7496	EIAPEEELAPBEEKVPVVAEVEEPPVPAVPEBPKKIIPEKKVEVINKPEAPPPEKPEP	7555
Qy	411	-----APTTPKE-----	417	Qy	876	DKTTER-DLRTTPTTTAAP-----KMTKETATTTTEKTTESKITATT	917
Db	6516	IHSITKREKEQVTEPAKVPKRVVABEKVPVPRKEVAPPVPRVPEVPEKELEPEEVAF	6575	Db	7556	EKVIEPKLKPDPPPPAPPEKEDVKEKIFOLKAIKKKVPENQVPEKV-----	7605
Qy	418	-----PTPTPKBPATTTKBPATTTKPE-----	439	Qy	918	QVTSTTTQDTPPKIT--TLKTTTTLAP--KVTTTKTKTTITTTIMNKPEETAKPK-----D	968
Db	6576	EEVVTHVEEYVIEEVEEYIIEEBEFITEEVEVPIPVKVEVPRKVPPEEKVPVPVKK	6635	Db	7606	-----ELTPLKVGGEKKVRKLLPERKEPEKEEVVLKSVLRKRPEEPEKVPKLE	7657
Qy	440	BPAPTA-----PKPK-----APTTPKEAPTTTPKBPATTTKPEPS--	474	Qy	969	RATNSKATTPKQKPTKAPKPTSTKPKTMRVRKPKTTTTPRKMSTMBELNPTSRIA	1028
Db	6636	KEAPPKAPVPEVPKPEEKVPVLLIPKKEKPPPAKVPPEVPEKVPV-BEKVPVVPVKVPEAPP	6694	Db	7658	KVKKPAVPEPPPKPEVEVEVPTVKRERKIPETK-----VPEIKPAI-----	7701
Qy	475	-----PTTPKEPAPTTTKSAPTTTKEPAPTTKPSAPTTPKPE-----PSPTTTKEPAPT	522	Qy	1029	EAMLQTTTRNQNTNSKLVENVPKSEADAGAGETPHMLLRPHVPMPEVTP-----	1079
Db	6695	AKVPEVPPKVPPEKVPVPAKKEVAPPA--KVPEVPKLIPPEEKPTFPVKKEVAPPK	6752	Db	7702	-----PLPAPEPK-----PKPE-----AEVKT-----IKPEVPEPEPTPIAAVPTVP	7739
Qy	523	TPK--EPAP-----	529	Qy	1080	-----DMDVLPRVNPQGIINPMLSDETNICNGKPVVDGLTTLRNGTLVAFRGHY	1128
Db	6753	VPKKEPVPVVALQEBEVLPEEBIVPEEVLPEEVLPEEVLPEEVLPEEVLPEEVLPEEVI	6812	Db	7740	VKKAEAKAPKEEAAPKPKIKGV-----	7763
Qy	530	-----TTPKBPAT-----	538	Qy	1129	FMMLSPFPSPSPS-----ARRITEVWIGIPSIDTV-PTRCNCEGKTFFKDSQV-----WR	1177
Db	6813	PPPEEVPPEEYVPEEVEFVPEEVLPEVKPKVPVPAVPEIKKVKVVKVVPKKEA	6872	Db	7764	-----PKXTSPSIEAERKLRPGSGGKPEPPEAPFT-----YOLKAVPLK	7803
Qy	539	-----TPKE-----PAPT-----	551	Qy	1178	FTNDIKDAGYKPIPIKFGGLGQIVA--ALSTAKYKNWPSVSVFFKRGSGIQOYIKQ	1234
Db				Db	7804	FVKEIKDILTESREVGSSAIFECULVSPSTAITT-----WMKDGSI-----RE	7847

R:Leimbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of *C. elegans* cosmid F35A5.
 A:Reference number: Z18485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB526
 A:Experimental source: strain Bristol N2; clone F35A5
 C:Genetics:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2

Query Match 14.1%; Score 1023.2; DB 2; Length 1274;
 Best Local Similarity 22.6%; Pred. No. 4.3e-14;
 Matches 354; Conservative 112; Mismatches 350; Indels 748; Gaps 82;

```

QY 77 KAPPPGASQTIKSTTKRSPKPNKKTKKVIESEITEEHSVSENSESSSSSSSSSS 136
D 3 RAPTP-----IKNPKW-KPP-----WESVDEEEMEVDDETPAPS 39
QY 137 TIWKIKSKNSAANRELQKLLKVDNKNRTKKPTPKPVVDVDRAGSLDNGDFKVTTPD 196
D 40 -----KLEKPSLKRKDAPTKV----- 57
QY 197 TSTTOHKNVSTPKITTAKEPI-NPRPSLPNSDTSKETSILVNKEITVEITNNKQT 255
D 58 -----PSGAPSPVFIKPP----- 71
QY 256 STDGKEKTTSAKETQSIETSA-----KDLAPTSKVLAK-----PTPKAETTTKGPA 304
D 72 -----VKWKAQWEDDEPMEAPAPVPAKKVDRDSPK-----KQPA-- 108
QY 305 TPKEPTT-----TPPKEPASTTPKEPTT-----TIKSAPTP----- 337
D 109 KPRDASPKKIMAAKBPETLPAVPTPVKNPVKKFKAPWEDDEVVDVDAFTVPAKKT 168
QY 338 -----KEPA-----PTTKSAPTTKPEP-----PTTP-KE 370
D 169 PVLKKEPAAAKPRDPSPKAAKKEHDIVPTPKNPAKWKAPWEDDEVDEYFTEEIKE 228
QY 371 PAPTTPKAPATTTKSAPTT-----KEPAPT-----TPKKAPATTTPKEPATTPEKPT 419
D 229 PEPATRKVPA--LKKEPSTSVKPSVDSPTKKVPVKKEPVP-----PTPKNPTKKWK 282
QY 420 -----PTTPKGPAPT--TKEPAPT--PKEPAPT--APKKPAPTTP----- 454
D 283 PWEDETPEVEVKPEPPVPEKKAPVLKKDPAPAAKARDPSKAAKPKKVPSSPVVPTPV 342
QY 455 -----KEPA-----PTTP-----KEPAPTTPKEPSTTPKEPAP----- 483
D 343 KNPVKYKPPWEVDEFAEENKPSAPEKKTPLVKKEPEPSSTTPSSDSPKAAPAVK 402
QY 484 -----TTKSA-----PTTTKEPA-----PTTKGAPTTP----- 508
D 403 PRDSSPKATPLQADPKAQEVPTPVKNPVKKYKPPWEVDEDPVEEVQPEAPAKKTPV 462
QY 509 -----KEPSPPTTKEPAPT--TPKGPAPTTPKKPAPTTPKE-----PAPTTP-KEPA-- 552
D 463 LKRKEPAAKDTAKPATSKTPTETPEKDPVKPRDSSPKVAAKPDQAAPATPVKNPVKKW 522
QY 553 -----PTTTKGPAPTAPKEPAPT-----TPKGPAPTTPKEEELAPITTPBP 577
D 523 RPWEDDETADDVSKPTDAKKTPLSLAKDPAPAKESLKPADTKAPAKPRDPSPKVAP 582
QY 578 TTPFKLITPT-----TPEKLPAP-----TPPKPAPTTPPEELAPITTPBP 615
D 583 TAPEKTPVLAKKEPAGPADSKTKEPEKSKPRDPSKAVPAKVPKT--EVAPAAVKKP 640
QY 616 TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEP-----APTTP-----KEPAPTTP 660

```

```

Db 641 EPIS--KPKDTAPKKAEPNSPVVP-PTPVKNPVKKKPPWEDDDAPAKPVSLPEPEKKTTP 697
QY 661 --KETAPTTPKG-----TAPTTLKEPAP--TTPKKPAP--ELAPTT 696
Db 698 VLAKKAPTDPDSEAAADPVSGPSSKDKLAKKAPVKPRDPSPMKAVPLKPAKTEVPFPAV 757
QY 697 TK-EPTSTTSD-----KPAPTTPKGTAPTTP-KEP----- 724
Db 758 VKKEPVAKGRDPSKKAKEPNSP--VVPPTPVKNPVKKKPPWEDDDAPAEPVNVPEP 815
QY 725 -----APTTPKEPA-----PTTPKGTAPTTLKEPAPTTP-PKKPAPKE----- 760
Db 816 EKKTPVLAKTPVKPRDPSKKAAPAKPTKTADAPPVSVKKPEVSKPKGESPKEAPENS 875
QY 761 --LAPTTTKGPT-----STTSKXPAPTTPKGTAPTTP-TPKEP-----AP--TTP 798
Db 876 PVVPPTPVKNPVKKKPPWEDDDDEPTBEVKKPSPEKKTPLVAKKEPEKPKDAPKVAAP 935
QY 799 KKPAP--TTPETTP-----PTTSEVSTPTTTTKBPTTHKSP--DEST 836
Db 936 RDPSPKKAPEKGPAPKVAAPKPRDLSPKKAIPIPANTQEAAPTVPKNPVKKKPPWEDDDE 995
QY 837 PELSAP-----TPKALENSPKPEPGVPTTK--TPAATKPEMTTTAKDKTTERDLRT 885
Db 996 P---AEPVSAPEPEKKTPLVAKKAPAKPRDPSKKAAPVAAKPD----- 1036
QY 886 TPETTAAAPKMTKE-----TATTTEKTTSKITATTQTSTTTQD 926
Db 1037 -PKIPEVPTPVKNPVKKKPPWEDDDDEPSEPVSAPEK----- 1075
QY 927 TTPFKITLTKTTTLAKVTTTKITTTTEIMNKPEETAKPKDRATNSKATTPKQKPTKA 986
Db 1076 -----KTPVLAKKAPTTPKATKPDSEAAADPVSGPTSKDKLKKAPVEKP-KPTTD 1125
QY 987 PK--KPTSTKPKTMRVRKEK-----TTPTRKMTSTM----- 1018
Db 1126 PKDDKLKSPAKKPEKAPAPAKKWKPVWDDDDPEADFTVPASKKPDDEDPADPLG 1185
QY 1019 -----PELNPTSRIAEAMLQTTTRENQFNKSLVEVNPKSEADAGAEGETPHMLLRPHV 1072
Db 1186 GPKTKDPLKNKKA-----PAEKTPKPKPEVSKPEPK----- 1218
QY 1073 FMEVPTPMDYLPVFNQGIINPMLSDETNCGKPVQGLTTLRNGTLVAFRGHFWML 1132
Db 1219 --PTEPPK-----PAAPKK-----WKPP-----WED 1237
QY 1133 SP-----FSPSPAPARITEVMGIPSPIDTVFRNCNCEGKTFPFKDSQYWRFTNDIKDAG 1186
Db 1238 DDEPEADFTMPAPKPDTE-----DPADSL-----G 1264
QY 1187 YPKP 1190
Db 1265 GKPK 1268

```

RESULT 7

T31108

cyst germination specific acidic repeat protein precursor - *Phytophthora infestans*
 C:Species: *Phytophthora infestans* (potato late blight agent)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #ext_change 09-Jul-2004
 C:Accession: T31108

R:Goenhardt, B.

submitted to the EMBL Data Library, April 1998

A:Reference number: Z20986

A:Accession: T31108

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1489 <GOE>

A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAAC7

C:Genetics:

A:Gene: car90

Query Match 14.0%; Score 1020.2; DB 2; Length 1489;

Best Local Similarity 24.9%; Pred. No. 7e-14;
Matches 431; Conservative 69; Mismatches 501; Indels 728; Gaps 79;

```

QY 27 LSKGRCFBSFGRBC-----DCDAQCKYD--KCCPDYESFCAEVH-----68
Db 43 IPCSGVGAEPV--GTACPKAGDVATSDCQPYLLSYNGAVCVAPVDAECALIHDDMWGCEP 100
QY 69 -PTSPSSKKA-----78
Db 101 PKTYTSVAEAEIAAYNGESSGWTGHDEVVQVGDEEBEIPARVNYDITVDTPIGVNC 160
QY 79 -----PPPSG-----ASQIKSTTKSRPKPNKTKKV 107
Db 161 VATETAQHATGGKYDTPSTGTQGDYGNTHYGSTTTEGVTGGYPTD-----AKV 216
QY 108 IESE-----111
Db 217 IDGETYLDYPTGITEIIEBDGTFGGYGTIDGGTTGGYTTVDNTHETEGAGGYDAGR 276
QY 112 -----EITEHVSVENQSSSSSSSSST-----TWIKSSKNSANRE 152
Db 277 EYESTTPTVGYSTETEGQHTGQYGFSDTEAFTEGTTVPRBEETAAFPSEDITYAPRE 336
QY 153 LQKKLVKONKKRRTKKKPTPKPPVDVDEAGSLDNG-----188
Db 337 V-----TPVAPTEKPPVDVEETTYVTEETTYAPTKSETNAPTERMHYAHIEKP 383
QY 189 -DFKVT-----TPDTSTTHQNKVSTSP-KITTAKPINPRPSLPNSDTSKETSUTVNKETT 242
Db 384 CDTEVIMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 438
QY 243 VETKETT-----TNKQSTDGKKTSAKETQSTKISAKDLAPTSLKVLAKGTPKAETTK 299
Db 439 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 495
QY 300 GPALTTPKEPTTPPKPEPASTPKPEPTTIKSAPTTPKEPAPT--TKKSAPTTPKEPAPT 358
Db 496 ASTEETTYAPTEETTYAPABETPYEPTPEET--TYAPTEETTYAPTEETTYAPTEETTYAPTE 554
QY 359 --TTKEPAPTTPKEPAPTTPKAPTTTKGAPT-----TPKEPAPTTPKAPAPT 406
Db 555 EEITYAPABETPYEPTPEETTYAPTEET--YAPTEETMYAPTEETTYAPTEETTYAPABET 613
QY 407 PKGAPPTPKPEPTTPKEP-----AP---TTKEPAPTTPKEP-----AP 443
Db 614 PYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTPEETTYAP 673
QY 444 TAPKAPAPTTPKBPAPTTPKEPAPT--TTKEPSPPTPKPEP-----APT--TTKSAPT--- 491
Db 674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTPEETTYAPTEETTYAPTEET 733
QY 492 -----TTKEP-----APT--TTKSAPT--TPKEPSPPTTKPEP-----APTPKBPAP 529
Db 734 MYAPTEETTYGPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 793
QY 530 TTPKBPAPTTPKEPAPT--TPKEPAPTTPKBPAPTAPKEP-----APTTPK 573
Db 794 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 853
QY 574 ETAPTTPKLTPTTPKEPAPT--TPKEP-----APTPEELAPT-----610
Db 854 TYAPTEKTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 913
QY 611 -----TPPEPTPTTPEEP-----APPTPKAAAPNTPKEPAPT--TPKEPAPT 651
Db 914 TEETTYAPABETPYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABET 973
QY 652 PKPEP-----APTPKETAPTTPKGTAP-----TTLKEPAPTTPKPEP-----AP 689
Db 974 PYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTPEETTYAP 1033
QY 690 KE---LAPT-----TTKEPTSTTSKPA-----PTTPKGTAPTTPKEPAPTTPK 730

```

```

Db 1034 TEETTYAPTEETTYASTEETTYAPTEETTYAPABETPYEPTPEETTYAPTEETTYAPTEET 1093
QY 731 EPAPPTPKGTAPT--TLKEPAPTTPKPEP-----APKE---LAPT-----TTKGP 769
Db 1094 TYAPTEETTYAPTEETTYAPABETPYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAP 1153
QY 770 TSTSDKAPPTTPKETAPT--TPKEPAPTTPKPEP-----APTTPETPTPTTSEVSTPT 820
Db 1154 TEET--YAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1211
QY 821 -TTKEPTTHKSPDESTBELSAEPTPKALENSKPEGVPPTTKTPAATKPEMTTAKDKTT 879
Db 1212 ETTYAPT--EETPYEPTPEETTYAPT-----EETTYEPTPEETTYAP-----T 1250
QY 880 ERDLRTPTTETTAAPKMTKET-----ATT--TEKTTESKIATTTQVTTST--TOD 926
Db 1251 EETTYAPTEETTYAP--TEETMYAPIDEETTYGTEETTYAPTEETTYAPTEETTYAPTEET 1308
QY 927 TT--PFKITT--LKTTTLAPKVTTTKTITTTIMNKPEETAAPKDRATNSKATTPKQP 981
Db 1309 TTYEPTGETTYAPTEETTYAPTEETTYAPM-----EET-----PY 1343
QY 982 KPTKAPKPTSTKPK--KTMPRVRKPKTTPTPRKMTSTMPBLNPTSRIAEAMLQTTTPNQ 1040
Db 1344 EPAEESTSTVSTKPCNTBEETDEPTDE-----PSDEPTD-----EPTDEPTD 1390
QY 1041 TPNSKLVEVNPXSDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPKVP--NOGIIINPML 1098
Db 1391 LPTE-----PSTPCDNQGI-----1405
QY 1099 SDETNICNGKPDGLATLRNGTLVAPRGHYFWMLSPPSPARRITEVWGI-----PSPI 1154
Db 1406 -----NLGIVENKYVYNA-----GLYNTTTPGP--1428
QY 1155 DTVTTRCNCSEGTFFPKDSQYWRFT-----NDIKDAGYKPIPKGFG--CLTGQIVAAALS 1207
Db 1429 -----RNSQSWHSCCRSCYND-----PICHAFSFHQTSSDSVCLETT 1464
QY 1208 TA-----KYKNWPESVYFFKRGSSIQYIYKQEPVQKCPGRRPALNYPVVGEMTQVRRR 1261
Db 1465 TSTSDREDOQNW-----LAGNMDR-----1484
QY 1262 RFERAIGPSOHTIRIOYSPARLAYQKGLVHNEKVSILWRGLPNVVT 1310
Db 1485 -----NVVT 1488

```

RESULT 8

S48478

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)

N;Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIRO19C

C;Species: Saccharomyces cerevisiae

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: S48478; A26877; B26877; S27281; JC6123

R;Rowley, K.

submitted to the EMBL Data Library, October 1994

A;Reference number: S48478

A;Molecule type: DNA

A;Residues: 1-1367 <ROW>

A;Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; G

J;Yamashita, I.; Nakamura, M.; Fukui, S.

J;Bacteriol. 169, 2142-2149, 1987

A;Title: Gene fusion is a possible mechanism underlying the evolution of STAL.

A;Reference number: A91831; MUID:87194600; PMID:3106330

A;Accession: A26877

A;Molecule type: DNA

A;Residues: 1-242 <YAM>

A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525

A;Accession: B26877

A;Molecule type: DNA

A;Residues: 762-1331 <YAM>

A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526

[illegible]

```

QY 786 -----APTPK----- 791
Db 2443 NPVGSKKRDVQLAVKKVGDAFTPAKNLEDRLITEGELTMDAKLNVKPKITWLKXDV 2502
QY 792 ----- 791
Db 2503 EITSDGHYKIVEEDGSLKLSILQTKLEDKGRITIKAESEBFGVAECASLGVVGRPKMAK 2562
QY 792 -----EPAPTT----- 797
Db 2563 PAFQSDIAPINLTGDTLECKLLITGDTFPVKWYIGTQLVCATEDTEISNANGVYTMKI 2622
QY 798 -----PKK--PAPTTPET----- 808
Db 2623 HGVTDMTGKIKVAYNKAGEVSTEGPLKVAPIPVFEFTSLCDATCREGDTLKLRAVLL 2682
QY 809 --PPPTTS-----EVSTPTTTKEPT----- 827
Db 2683 GEPEPVSVYVNGKLEESQNIKHSEKGTVTVIKDIICDYSQVVCEALNEVYKATSE 2742
QY 828 -----IHK----- 830
Db 2743 ATLLVLPGEPPDFLEWLSNVRARTGTVVHKVVFVTDGPKPSLTWYINNKEILNSDLYTI 2802
QY 831 -----SPD-----ESTPELSAE-- 842
Db 2803 VTDDKTSTLTINSFNPDPVHVGELICKAENDAGEVSCVANMITYTSDMFSESESAQAEEF 2862
QY 843 -----PTPKALENSPK-----EPGVPTTK----- 861
Db 2863 VGDLDTEDESLREMHRTPTPVM--APKPTIKDKTKAKKGHSAVFECVVPDTKGVCCK 2919
QY 862 -----TPAAT-- 866
Db 2920 WLKDGKEIELIARVQTRTGPPEGHITQELVLNDVTPEDAGKYTCIVENTAGKDTCEATL 2979
QY 867 -----KPMWTTTAKDKTTERDLRTT-----PETTTAAPKMTKETATT 903
Db 2980 TWIESLEKKSEKAPFEIVALQDKTKTSEKVLVECKVGEKPKVSWLHNDVSRKNPN 3039
QY 904 TEKT-TESKITANTQ--VTSSTTQDTPPKITTL-----KTTTLAPKVTTTKT----- 950
Db 3040 SEKITQESITVBSVEGVERVITTSB-----LSHGQKYTCIAENTEGTSKTEAPL 3090
QY 951 -----ITTTETMKPK-----BETAKPKDR----- 969
Db 3091 TVQGEAPVFTKELQNKELSIGELVLSCSVKGSQPQPHVDFYSFSETTKVETKITSSRIA 3150
QY 970 -----ATNSKATTPKPKPT--KAP-----KKPTST 993
Db 3151 IEHDQNTHRWVISQITKEDIVSYKAIATNSIGTATSTSKITTKVEAPVFEQGLKKTSV 3210
QY 994 K-----KP-----KTMPR-----VRKPKTT----- 1008
Db 3211 KEKEIKMEVKVGGAPDVEWFKDDKPVSEDDGNHEMKKNPETGVTFLVVKQAATTIDAGKY 3270
QY 1009 -----PT-PRKWTISTWPELNPTSRIAEAMLOTTTR----- 1037
Db 3271 TAKASNPACTAESSAAEVTSQLEKPTFVRELVTTEVKINET-----ATLSVTVKGVDP 3325
QY 1038 -----PNQTPNSKLV-----EVNPKSEDAGGAG 1061
Db 3326 SVEWLKDGQPVQTDSSHVIAKVEGSGSYITIKDARLEDGKVCATNP-----AGEAKT 3381
QY 1062 ETPHMLLR---PHVFMPEVTP-----DMD----- 1082
Db 3382 EANFAVKNLVPPEFEKLSPLEVKESTTILSVKVGVPFSPVSEWFKDDTTPISIDNVHV 3441
QY 1083 -----YLPVPVNO-----GII-----INPM 1097
Db 3442 IOKTAVGFSLTINDAQGDVGIVISCRARNEAGEALTATNFGLINDSIPPEFTQKLRLP 3501

```

RESULT 10

T34513

hypotheical protein ZK783.1 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

```

QY 1098 -----LSDETNICNKPVD-----GLTTLRNGT 1120
Db 3502 EVREQETDLKLVTVIGTPVENVFKDD-----KPINIDNSHIFAKDEGSGHLL--T 3552
QY 1121 LVAFRGRH-----YFWMLSPF--SPSPARRITE 1146
Db 3553 IKQARGEDGVYTCATNEAGEAKTTANMAVQBEIEAPLFVQGLKPYEVBQGPABLVR 3612
QY 1147 VWGIPSP----- 1154
Db 3613 VEGKPEPEVWFKDGVPIAIDNQHVIEKKGENSHLTVIKDTNNADFGKYTCQATNKAGK 3672
QY 1155 -----DTVFTRCNCEGKTF-----PFKD 1172
Db 3673 DETVGLKIPKYSFEKQTAEEVKPLFTEPLKETFAVEGDTVWLECKVNKESHPQIKFFKN 3732
QY 1173 SQYWR-----TN-DIK----- 1183
Db 3733 DQPVETGOHMQLEVEDGNIKLTIQNAKEDVGAYRCEAVNVAGKANTNADLKIQFAKV 3792
QY 1184 -----DAGYKPKPIF----- 1192
Db 3793 EEHVTDESGQLEBIGOFETVGDTPASSKTDTGRGAPFVELLRSCVTTEKQQAIIKCKVK 3852
QY 1193 -----KGFGLT----- 1199
Db 3853 EPRPKIKWTKEGKEVMSARVRAEHKDDGTLTLTFDNVTQADAGEYRCEAENEYGSATWE 3912
QY 1200 GQIVAAALSTA-----KYKNWPE-SVYFFFRGGSIQ- 1228
Db 3913 GPIIVILEGAPKIDGAPDPLQPVKPAVVTVGETAVLEGKISGKPKSPKWKYNGBELPK 3972
QY 1229 -----QY-----YKOBPVQKCFGRPP 1245
Db 3973 SDRVKIENLDDGTQRLTVTNAKLDDMDREYRCEASNEFGDVMSDVTLITKEPAVAG- 4029
QY 1246 ALNYPVYGEWTVRRRRPFA----- 1266
Db 4030 -----PFKLSAIQVKTETETAKFECKVSGTKPDVKWFKDGTPLKEDKRVHFESTDDGTQ 4084
QY 1267 --IGPSQTH--TIRIQSPARLAYODKGVLLHNEVKVSI-----LWGLPNV----- 1308
Db 4085 LVIEDSKTDDQGNRYLEVS-----NDAGVANSKVPPLTVVPSETLKKIKGLTDVNVTOGT 4138
QY 1309 -----VTSA----- 1312
Db 4139 KILLSVEGKPKTVKWKYKGTETVTSSTQTKIVQVTESEYKLEIESAEMSDDTCAYRVLS 4198
QY 1313 -----ISLPNIRK----- 1320
Db 4199 TDSFVSSESATVTVTKAAEKISLPSFKGLADQSVFKGTPLVLEVEIEGPKDKVWKYKNG 4258
QY 1321 -----PD----- 1322
Db 4259 DEIKQKVEDLNGKXRLTIPFOEKDVGESYVTAANEAGEIESKAKVNVSAKPEIVSGL 4318
QY 1323 -----GYDYAFSK-----DOYINIDVP----- 1340
Db 4319 VPTTVKQGETATFNKVKGVKWKYKNGKEIPDAKTKDNGDGGSYSLIENPAQVEDAAD 4378
QY 1341 -----SRTARAI-----TTRSGQTLIS--KV----- 1358
Db 4379 YKVVVNSDAGDADSAALTVKLADGDKQVKPEIVSGLIPTTVKQGETATFNKVKGPVK 4438
QY 1359 --WY 1360----- 1360
Db 4439 QVKWY 4443-----

```

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34513
R:Pavello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 13.3%; Score 965.1; DB 2; Length 3507;
Best Local Similarity 14.5%; Pred. No. 6.6e-12;
Matches 453; Conservative 199; Mismatches 512; Indels 1959; Gaps 108;
QY 19 IQQVSOELSCKRCFSFE-RGRECDCA-----QCKYDKC----- 55
DB 506 VELTISGLACTSYCPNSECYGCYCVSGYGNALVGCEDIDECITEICNIEANWCV 565
QY 56 -----CPDY----- 59
DB 566 NLIGGFVCCNPTNATHDDCIDFLT VKIYAMIIIFLLKGLIETKXGLHVIIGNEDTV 625
QY 60 -----ESFCAEVH----- 67
DB 626 VATRSNHSSTDQLITVQVSRNPFSTGQIILTRCKVSSGEAVTQTTDADEFLGLISAADL 685
QY 68 -----NPTGPPSKKAPPSPGA-----SOTIKST-- 91
DB 686 AGSSGILPTLPKLEGSKKA---SGGVWHEDEGEDEDLMEEGSGSWITTINGI 742
QY 92 -----TKRSPKPNKKTKKVIIS-----BEITEHSVSEN 122
DB 743 TGSPRSEGTIRVITLGEDPEPATKPGISAPDKTGEKSTESDGEELITVEKDGEA 802
QY 123 QESSSSSSSS----- 132
DB 803 QSSGSSATSGKKSEATSGSSSSSAKSGTGEAGSSGASSSSGSVSGSVSTESG 862
QY 133 -----SSSSTIWKIKSKN----- 146
DB 863 SFGTSSSGSVSEATGTVGDSGSKPSKSTEEKLPFTKNGEKSPISGSDTTGKESSE 922
QY 147 -----SAANRELQKKLV----- 159
DB 923 ETTGRKPIEGDSLTEGSGGSEWFTGSKGHFSGSKSVTSKGPTQSGAEGSGSGPKV 982
QY 160 -----KXNK----- 163
DB 983 PKGCPAPIITDGEBSSTSTGDKSGKPKADKSNKVPKTKDNKPOLITDGEDSTSETS 1042
QY 164 -----KNRTK-----KQTP-----RPPVVDEAGSLDN 187
DB 1043 GGEQGPKSKGQPPGDKGSEVKKPTSEVDGPGNLSGTGKSNVPLKPTDLPEEGSGI-- 1100
QY 188 GDFKVTTD-----TSTTQH-----NKVSTSPKITTAKINPRPSPLPNSDTSKE 232
DB 1101 -----LTTSSGKNSTFEHGTKLERLPKPTEDKSETPOLGLEISAGKKPE--PEDGTSKE 1154
QY 233 TSLTVNKETTVETKETTNNKQT-----STDGKEKTTISAKET 269
DB 1155 VGLLEIWEEST--TPGSTILDSVGLIEISGDLTKATKKPHVELEGSGTGDEIAT--T 1210
QY 270 QSIKETSIAK-----DLAPTSKVIAKPTPAETTKGPAITTPKPT----- 310
DB 1211 RDVSKSTKKPRVEVDGNGNETSGVDGKPTTPAETPSSAESSTRIPTTSEASPEGSG 1270

QY 311 -----PTTPK---EPASTTKPEPTTIKSAPTTKPEAPTITTKSAPTTP-----KEPAPTT 359
DB 1271 EAGVPESPDGSGESSTAPDGVSPTSATAPEVP-----TTSASTPDAVESGIPST 1323
QY 360 TK---EPAPTTTKPEAPTTP-----TKBPAPTTTKSAPTTPKPEAPTTP 399
DB 1324 SKPTAEPLETT---APSTEIVTSPEGSGETEESLPTPTEGSGESTTSSAPTIV--EPATVLP 1377
QY 400 -----KKPAPTTTKPEAPTTP----- 414
DB 1378 PQNRNEKPEPTKDTFALPTTTTGAPOANDSVVENTKCTSSDECGLDLCERRTGVCRCEP 1437
QY 415 -----PKE-----PT----- 419
DB 1438 GFEGAPPKKSCVDVDECATGDHNCHESSARCONVVGACFCPTGFRKADGSCQDIDECT 1497
QY 420 -----PTTPKPEAPTTPK----- 431
DB 1498 EHNSTCGANAKCVNKPVTGTCBENGFLGDGYQCVPTT-KKPCDSTQSSKSHCESNMS 1556
QY 432 -----EPAP----- 435
DB 1557 CEYDITVDGSVECKECMGYKSGKVCEDINECVAEKAPCSLNANCVNMNMTFSCSCKQY 1616
QY 436 -----TTTPKE----- 440
DB 1617 RGDGFMCTDINECDERHPCHPACTNLEGSKFCECHSGEGDGIKCTNPLERSCEDE 1676
QY 441 -----PAPTA----- 445
DB 1677 KFCGRVDHVSLSVRIYNGSLSSVCECEPGFRPEKESNCVDIDECEERNNDPASAVC 1736
QY 446 ----- 445
DB 1737 VNTGSYRCECAEGYEGEGVCTDIDECRGMAGCDSMAMCINRMGCGCKMAGYTGDG 1796
QY 446 -----PK-----KDPATTPKEPAP----- 459
DB 1797 ATCKIEEPKSDKTACTDEWSRLCELEKQCTVDEEVQCGACLPGHHPINGTCOSIQ 1856
QY 460 ----- 459
DB 1857 ISGLCAQKNDCKNHAECIDIHPDSHFCSCPDGFDGMIODVDNAGMAGCDDENTKCE 1916
QY 460 -----TTTPKE-----APTTPKPSPTTPKEPAPT 484
DB 1917 NTIGSFNCVLEGFKKVDKCVDEKKQPNREKIEIDEENSSSSSSQEKPTTKGIVSST 1976
QY 485 TTKSAPTTPKPEAPTITTKSAPTTPKPEPSPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEPA 544
DB 1977 SATSSESTTAEPHVTTSSISSTTKDMTSSKSPENVTMSSSESPEVSTSSSKSTTASETV 2036
QY 545 PTPKPEAPTTPKPEAPTAPKPEAPTTPKETAPTTPKLLPTTPE-----KLAPTT 595
DB 2037 SSTPSSESS-----SSEAPLITSSPATTTTETV--ITESSVSKSTTPKESSSEITVKLSKS 2087
QY 596 PE-----KPAPTTPEELAPTTPEEPTTPPEEAPTTPKAAAPNTKPE-----APT 643
DB 2088 PEVTESSVKSSPSPS-----TTSQSVTSTVPETSKSTVLSSEAPVTSPTSEVTSSETK 2143
QY 644 PKEPAPTTPKPEAPTTPKET-----APTTPKGT-----APTTPKPEAPTTPKPEAPKELA 693
DB 2144 PLSASSTTGTNTTSTSSLASVKSTSAPEGTSASVAPVKLSLSPDV--SQSPSTKTFD 2202
QY 694 PT-----TTKEP-----TSTTSKDPAPTTPKGTAPTTPKPEAPTTPK 730
DB 2203 ATESSTVQASSTSGTSVKSTSEPESHVTKLSITSSNPSPPSVPTSPKSTPTVPEST--E 2260
QY 731 EPAPTTPKG-----TAPTTPKPEAPT--TPKKP 756
DB 2261 QPTSTTTFSGQLTPMNSNSELVITSEBHVLSLSLSPDVQSQSTTFNNLSESTVETPKTS 2320

```

QY 757 APKEL---APTITKGPST---STT-----SDKP-----777
Db 2321 SEVLSNSEPSTTEAPTTLSPDILSTTTNNLSQSSTVSTEDSEISENSEKPTSAPELV 2380
QY 778 -----APTTP-----KETAPTP-----KEPAPTT 797
Db 2381 TSSVTHVASSGPDVPTESSEPDLLTGSSTENIPEASSKQTISSTPTDPTTASSEPTKST 2440
QY 798 PKKP-----APTTP-----807
Db 2441 SMSFDLSTSNVLSSESTTPSSSKSPVSSSTEGISVVSTFEFKVPSTISSVLEEDLT 2500
QY 808 --TPPTTSVSTPTTKEPT-----IHK--SPDESTPELSAPRTPKALENSP-K 853
Db 2501 KTTSPILEETTTASSETSEPLETSLTVSVRIHELTTSSSENVKPESESTTSSSSKPSQ 2560
QY 854 EPG-----VPTT-----KTPAATKPEM-----TTTAK 875
Db 2561 EPAGILSTVVVPTSSVSLITASEIAITSNTPFKQGRTPITTSFKSLVKSTTSPTSSTVS 2620
QY 876 DKITERDLRTTPETTTAAPKMTKETATT-----TEKTESKITATTTQVTS-- 921
Db 2621 SEPESEKRTIVSTVSTTPTTEFTTSSSLLITAAFSKPTSTESSEAPTTPAKTSET 2680
QY 922 -----TTTQDTPPKITTLKTT-----TLAPKVTTTKKTIITT-----954
Db 2681 KPSNVSTSRKSTENVETSTSQSGLESSTMSSTSSRPEINAPATVVSSEASSTLEENS 2740
QY 955 -----EIMNKPEE-----962
Db 2741 STSPTSSEASVKLSLFPESITSEAVTVSSRAPEITMSSESHREISTVSSEPESEIP 2800
QY 963 -----TAKPKD-----RATNSK 974
Db 2801 LSTTVSNVVVTASSIPSEEPILSVSTSSSTPRVRLITGTDDLLIVSTVSHGNRRNIT 2860
QY 975 A-----TTPKQKPTKAPKPTSTPKPKTMFRVRKPK---TTPTP-- 1011
Db 2861 ASSVPSNSTSPIILPSESLLTPQPPPTTTTAKPATTSGRGPPSIQPPAEMTTTAPPP 2920
QY 1012 -----RKMTSMPELNTFSRIAEMLQTTT-----1036
Db 2921 PSNGGYGEETNOBEEQVTS-----TTTEAPSLCSTVTCHSLATCEQSTGVCICRDGFI 2974
QY 1037 -----RPNQTNESKLIVEPNPKSEADAGAGEHTPHMLR 1069
Db 2975 GDGTTACKSKSTADCLSLPSLCADKAKCNSTRSCEC-----DAG-----3014
QY 1070 PHVFMPEVTPDMDYLPRVNPQGIINPMLSD---ETNINCKRPV-----DG 1112
Db 3015 -----YI-----GDGYVCSPHQDCVLRDLNLSPEAVCQNRRCQCLPGFTGDG 3057
QY 1113 LTTLR-----NGTLVAFRGHYFWMLSPLFPSPSPARRITEVWG-----1149
Db 3058 VKCVSITHERASNCSCDANNAHCVGGTCKCNPGYF-----GNGLCC 3098
QY 1150 IPSPIDTV-FT-----RCNCEGKTFFFFKDSQYWFRTINDIKDAGYKPIFG 1194
Db 3099 VFDELDCVHTGTICHPNVNCNSESRCQCS-----SG 3130
QY 1195 FGG-----LTGQIVAA-----LSTAKYKNWPESVYFFFKGSGSIQQYVYKQEPVQ 1238
Db 3131 FSGNGVSCFPKSCRTDKSVCAKNAICLPT-----GSCICRHGFKGDPFY 3175
QY 1239 KCP---GREPA-----LNPVYVGEWTVQVRRRPER---ATGPSQT--- 1272
Db 3176 KCSSLVSKEPANQODLSVDSSCVTPCDSTQLCISGEC---ICKSGFRNRNSTLSGSETCAD 3233
QY 1273 -----HTIRIQYSPARLAVQDKGV 1291
Db 3234 IDECAEKSHKDRVATCRNTFGSHVCTCPDGHVGDIITCVPHV-----3276
QY 1292 LHNEVKVSILWR--GL-----PNVVTSAISLPIRKP-----1321

```

```

Db 3277 --NOKUSVYCEADGTMILVLGNETADFGKIFVKGQAEINPYCSKSFSSLLNSRKPYFVKV 3334
QY 1322 -----DGVDY-----1326
Db 3335 VFQCHDVQLLDNHTWASTVVVVKHAMFLTNKADSYDLRCQYPIGSRVAVSHVNVSELAATT 3394
QY 1327 -----YAFSKDQYYNID-----VPS-----RTARAITTR 1350
Db 3395 STLTEKNSTLAPICRLSVSNDOHSSISSAMVGVDTLKLALEVTPSENFGLPRNCFAINIE 3454
QY 1351 SGQ 1353
Db 3455 SGE 3457

RESULT 11
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-May-1992 #sequence, revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: ventral prostate
A:Note: sequence inconsistent with the nucleotide translation
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
Submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:g206397; PIDN:AAA41950.1; PID:g554494
R:Peeters, B.; Heyns, W.; Bosseyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary s
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PEB>
A:Note: this peptide, designated proline-rich polypeptide V, can be found at several loc
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
A:Note: this peptide, designated proline-rich polypeptide IV, can be found at several lo
C:Genetics:
A:Introns: #status absent
A:Note: single copy gene with no introns
C:Superfamily: rat prostatic proline-rich peptides 637K precursor
C:Keywords: prostate; tandem repeat
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MA

Query Match 12.6%; Score 919.5; DB 2; Length 5762;
Best Local Similarity 11.4%; Pred No. 1.8e-10;
Matches 526; Conservative 196; Mismatches 474; Indels 3404; Gaps 150;

```

```
QY 35 ESFERGECDDAQKKYDKCCPDY-----ESFCAEVH-NFTSPSSKKAPP 80
Db 850 ETTYNKQVDSLSTHLKTSGLPDYTWGLEPSLYQQITQSPPKSLHENPKSP-----902
QY 81 PSGASITKSTKSPKPNKKTKKV-----TESEITEEHSVSEN 122
Db 903 -----VLKPSAAQALKPRKEKLSPTNNMI PHHPKPKLKNMWTTHIPAHKMTVPRQIQED 956
QY 123 QESSSSSSSSSS-----SSSTIWIKIS-SKNSAANRELQKLVKDNKNKRTKK-- 169
Db 957 QGYTITSSNGSFQPLDLVLTSGIIPVVKHILPKRTVNPQTVSQVKISHSQHVETQHPN 1016
QY 170 -----KPTPKPPVDEAGSLDNGDFKVTTPDTST-----199
Db 1017 SETTVQPLDLFAINLQPTPKE-----NFAQTFQDTTQMIGPCKEVIAQA 1063
QY 200 -----TOH 202
Db 1064 PEHHEGTIPIDQABYSTLPVTSFQPLDQELTITSEAIREFPHPTVPOQTIIIVHPEH 1123
QY 203 NKVSTSPKITAKP-----INRPS-----222
Db 1124 PLVHSEQVYTOHPNTEFALIQPLDLELITPQPTABGELPQTLQDSTTQIIEPPTVVVG 1183
QY 223 -----LPPN--SDTSKETSIVNKETTVET-----KETTTN-- 252
Db 1184 PVPIYEVTVQTTSDQAEYPPSPVTSFQSLDLELITSPATRESYHPSLLQOQTIVNPP 1243
QY 253 -----KOTSTD-----GREKTTSAKETQIEKTSKADLAPTSKV 286
Db 1244 EHELVIHSEQVHTQHPNLTEATVQPLDPLDITPQPTTEGELPQTLQDS-----TSQI 1296
QY 287 LAKPTPK-----AETTTKGPAITTP-----KEP-TPTTPKEPA 318
Db 1297 IIEPPIKVALPVYQVSDQAEYTTSTVSFQPLDQELTITSEAIREFPHPTVPOQTI 1356
QY 319 STTP-----KEPTPTTKSAP-----TTPKEPA-----PTTKSAPITTPKEPA 356
Db 1357 TVHPTKPLVTHSRQTOHPNTEVTVQPLDLELMTPTPTQPTABGELPQTLQDSTTQIIEP- 1415
QY 357 PTITKEPA-----TTPKE-----PAPT-----375
Db 1416 PTVVVGVPYIEEVTVQTTSDQAEYPPSPVTSFQSLDLELITPTEPEHEITQKTIVFP 1475
QY 376 -----TKBPAPT-----TTPKAPT-----394
Db 1476 PMYTDVTLPOQVSVQHLKPTBEGIVQPLDLELITPQPTTEGELSQTVQESTQNKEPHKE 1535
QY 395 --APT-----TPKK-----401
Db 1536 VVAPVPVYQAVTVPTFSQYQAEYKSLQPLDLELTVTSEPTKEAVHSTISKNSLAINQY 1595
QY 402 -----PAPT-----PKE-----409
Db 1596 VHIQHPNPAEATVQPLDLELITSSSQPTABGELLYSMQEVTVQISSEPPKQVWTVPVPEYQ 1655
QY 410 -----PAP-----TTPKEPTPTTPKE-----425
Db 1656 EVAVPAPYQDAKYPLSSIVSLNSLDQELTILSSELLGEAHQITTPDETWVLPPKDRQGIY 1715
QY 426 -----425
Db 1716 PDHDHKLNLTEVNTQPHLEHTVQHOFTIEBERSQSIQKKTQITQITEPGKKVQVLAQES 1775
QY 426 -----PAPTKEPAPTTP-----438
Db 1776 EEVTPMELIKETAPFTPHSMALQSLDEKLTIHSHSPGWTQOQHANKESKGTHTGKILLD 1835
QY 439 -----438
Db 1836 YAEPNMEIELKHGLFLKTTBATESNTQMTKSLKQVTLFTQNKKSMLPALVESQD 1895
QY 439 --KEPAPTAPKP-----APTTPK-----EPAPT-----460
Db 1895 -----460
QY 1896 ESQPPNMSLQPLDQELTSSQPHGWPVHPNTPKIXYLHYAEP-PTGPFVEPPELFFLK 1954
QY 461 -----TPKE-----PAP-----TT 469
Db 1955 TTKSRPQVQTATQMAASPKEWMSRAPENKEAVLSGPGEDQDESPSPNMSLQSDOELTL 2014
QY 470 TKEPSPTTTPKEP-----APT-----TTKSAP-----490
Db 2015 SSOPHGWIPHPNTHGKIYLYHYAEPPTGFVEPDPDLFFLKTYSKPVWVWTLTRTKSRKE 2074
QY 491 -----TTTKEP-----496
Db 2075 MVSQSPKYEEAVLVHGEQESRSPNMSLQSLQELTSSQPHGWPVHPNTHGKIYL 2134
QY 497 --APT-----TTKSAP-----505
Db 2135 HYAEPPTGFVEPDPDLFFLRTTKSKPVQGTATRMVKSPEEMVSLDPENKEAVTFAQGEK 2194
QY 506 -----505
Db 2195 GESPPSPNMSLQSLDHELFMSQPHGWI PHHPKTPDKIYLYHYAEPPTGFVEPDPDLFFLR 2254
QY 506 TTPKEPSPTTTPKEPAPTTP-----KKPAP-----537
Db 2255 TTKSKPVQGTTEMA-KSPKEMVSOQTPYKEAVLSGPGEDQDESPSPNTSLKSLDOEVA 2313
QY 538 --TTPKEPAPTTPK-----EPAP-----TTPK-KPAPTAP-----564
Db 2314 MSSQPHSGVPHHPKTPGKIYLSIEPPPGPVFKPTDILVKTITTKSKPAEWTPRIDKLL 2373
QY 565 KEPAPTTPK-ETA-----TTPKKL 583
Db 2374 KEMVPHSPEYEEAVFPAGHGQDESGSPNMPLOPLDQELTSSQPHGWPVHPNTHGTEGKI 2433
QY 584 -----TPTTPEKLIAP-----TTPKEPAPTPEELAPTTPBEPTTPP-----620
Db 2434 YLYHYAEPPTGFVEPDPDLFFLRTTKSKPVQSPKXIA-KSPKEMVSOQTPYKEAVLSGPG 2492
QY 621 --EBAP-----TTPKAAAPNTPK-----EPAP-----641
Db 2493 EDQDESPSPNMSLKSLDQEVMTSSQPHSGVPHHPKTPGKIYLSIEPPPGPVFKPTDLI 2552
QY 642 --TTPKEPAPTTP-----KEPAPTTPK-ETA-----664
Db 2553 LVKTTTKSKPAEWTTPRIDKLLKEMVPHSPEYEEAVFPAGHGQDESGSPNMPLOPLDQ 2612
QY 665 -----PTTP-----KGTAPTTL- 676
Db 2613 ELTSSQPHGWPVHPNTPGKIYLYHYAEPPTGFVEPDPDLFFLRTPKSKPVQGT-PTQMA 2671
QY 677 -----KEPAP-----681
Db 2672 KSPEEMVSLSPKNKETVFPAGQKQDESPSPNMSLQSDOELTSSQPHGWI PHHPNTH 2731
QY 682 -----TTPPKP-----APKEL-----692
Db 2732 GKILYHYAEPPTGFVEPDPDLFFLKTYSKPVQSPROIDKSKPEMFTQSPYEESLLPA 2791
QY 693 --AP-----694
Db 2792 HAEGQESRAPPHFSLQPLDQELSLSHPHGWI PHHPNTPDKIYLYHYAEPPTGFVEPDP 2851
QY 695 -----TTPTEP-----TSTTSK- 710
Db 2852 LFFLKTYSKPVQGTATKIDKSPEDRVSOQTPYKEAVLSGPGEDQDESPSPNMSLQSD 2911
QY 711 -----TTPKGTAPTTPKEP-----APT-----727
Db 2912 QELATSSQPHGWI PHSPNAPDKIYLYHYAEPPTGFVEPDPDLFFLKTYSKPVQGTQMA 2971
QY 728 --TTPKEPAPTTPK-----738
```


Db 2972 KSPKEMVOTPEYKEADLSAPGENRDESPSSPNMSLHPDQELSLSSQPHGWIPHPNTH 3031
Qy 739 -----GTAP-----TTLKEP----- 748
Db 3032 GKILYHYAEPPTGPFVEPPDLFFLKTTKSPVQSGPSQIDKSPKEVFSQSESESVLPA 3091
Qy 749 -----AP----- 750
Db 3092 QAEGQESRAPPHMSLQDQLDQKLTLSHSHPGWIPHPNTPGKIYLYHYAEPPTGPFVEPPD 3151
Qy 751 -----TTPKXP-----APKEL-----APTTKG----- 768
Db 3152 LFFLKTTKSPVQSGPSQIDKSPKEVFSQSESESVLPAQAEGQESRAPPHMSLQD 3211
Qy 769 -----PTS-----TTSKAPATP----- 782
Db 3212 QDLTSSHPHGWITHHNTPDKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPAQWTTQMA 3271
Qy 783 -----KETA----- 786
Db 3272 KSPPEMVSLSPENKETVPPAQGQGESISPPHMSLQDQLDLPSSHHPHGWIPHPNTH 3331
Qy 787 -----PT-----TPKE-----PA 794
Db 3332 GKILYHYAEPPTGPFVEPPDLFFLKTTKSPVQSPRQIDKSPKEVFTQSPYEESVLPA 3391
Qy 795 ----- 794
Db 3392 QAEGQESRAPPHMSLQDQLDQLTSSHPHGWITHHNTHGKIYLYHYAEPPTGPFVEPPD 3451
Qy 795 -----PTTPKXPA----- 802
Db 3452 LFFLKTTKSPVHSGPRQIDKSHKEMFTQSPYEESVLPAQAEGQESRAPPHMSLQD 3511
Qy 803 -----PTTPTP-----PPTSEVSTP-----TTTK-----EPTTIH 829
Db 3512 QDLTSSHPHGWITHHNTPDKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPVQSGPRQID 3571
Qy 830 KSPDE-----STPELSAEP----- 843
Db 3572 KSPKEMFTQSPYEESVLPAQAEDQESRAPPHMSLQDQLDQLTSSHPHGWITHHNTP 3631
Qy 844 -----TPKALENSPKX----- 854
Db 3632 DKILYHYAEPPTGPFVEPPDLFFLKTTKSPVHSGPRQIDKSPKEVFTQSPYEESVLPA 3691
Qy 855 -----PGVP----- 858
Db 3692 QAEDQESRAPPHMSLQDQLDQLTSSHPHGWITHHNTPDKIYLYHYAEPPTGPFVEPPD 3751
Qy 859 -----TTKT-----PA----- 864
Db 3752 LFFLKTTKSPVHSGPRQIDKSPKEVFTQSPYEESVLPAQAEDQESRAPPHMSLQD 3811
Qy 865 -----ATKP-BMTTAKD 876
Db 3812 QDLTSSHPHGWITHHNTHGKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPVQWTTQID 3871
Qy 877 KITERDLRTPE----- 888
Db 3872 KSREEMVSPENEEADIPRHGQDQLRNPIISLQDQLTSSQPHGWVPHPPSTP 3931
Qy 889 -----TTTAAKMTKETATTE----- 905
Db 3932 DKILYHYAEPPTGPFVEPPDLFFLKTTKSPVQGTPTQIAKSPKEMVFTQPEYKAULSA 3991
Qy 906 ----- 905
Db 3992 PGEDQDPTSSPNMSLQDQLTSSQPHGWITHHNTHGKIYLYHYADPTGPFVEPPD 4051
Qy 906 -----KTTESK-IPATTQVTSITTO----- 925
Db 4052 LFFLKTTKSPVQGTPTQMDKSPKQGLLPAHGEAQDESPPNMSLQDQLTSSQPH 4111

Qy 926 -----DTTPFKITT-----LXTT----- 938
Db 4112 GMITHPNTPGKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPKLAGTPTQMAKSPKEMFSQT 4171
Qy 939 -----TLAP----- 942
Db 4172 LEHKEAVLSAPGEQODESPSSSNMSLQDQDITLSPOQYGRIPHPVTPSGKIYLYHYVEH 4231
Qy 943 -----KVTITTK-TITTTIMNKPEE----- 962
Db 4232 PTVPLFQHPNMFSTKTTTISKPLQLTQMAKSPBEIVSLRPEYKEAILPAQVESOESESVF 4291
Qy 963 -----TAKPK 967
Db 4292 LPNMSLQDQLTSSQPHGWIPHPNTPDKIYLYHYAEPPTGPFVEPPDLFFLKTTKSK 4351
Qy 968 D----- 968
Db 4352 PVQWTTQIDKSLKQMAQSPYEAEVPAHGEQDESRYAPNISLOSLDHDTLSSQPH 4411
Qy 969 -----RATNSKAT--TPKP----- 980
Db 4412 GMITHPNTPDKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPAQWTTPTMAKSPPEMVSL 4471
Qy 981 -----QK 982
Db 4472 LKNKEAVLPAQGDDEESISPPNMSLQDQLTSSQPHGWITHHNTHGKIYLYHYAEP 4531
Qy 983 PT-----KAPKK----- 989
Db 4532 PTGPFVEPPDLFFLKTTKSPVQGTITQMVKSPKQGLLPAHGEAQDESPPNMSLQD 4591
Qy 990 -----PTST-----KKPKTPRVRKP-----KT-----TPT----- 1010
Db 4592 QELSLSQPHGWITHPNTPDKIYLYHYAEPPTGPFVEPPDLFFLKTTKSKPLLTPTQMA 4651
Qy 1011 --PRKMTSTPELNPSTRIAEAMLQTTTRNQT-----PNSKL----- 1046
Db 4652 KSPKEMVOTPEYKEAVLSGEDLDESTSSPNMSLQDQLDQLTSSQPHGRIPHVTSPGK 4711
Qy 1047 -----VEVNPKS-----EDAG 1057
Db 4712 IYLYHYAEPPTGPFVEPPDLFFLKTTKAPVQGTATHVEKSPKEMVFRSPYKQAVLPAH 4771
Qy 1058 GAEGET--PHMLL-----RPHVEMPE-----VTPDMD 1082
Db 4772 EAQDESPPNMSLQDQLDQLTSSQPHGWIPHHNTHGKIYLYHYAEPPTGPFVEPPDLF 4831
Qy 1083 YLPR-----VPNQ-----IINPM----- 1097
Db 4832 FLKTTKSPVQGTATQMAKSPBEIVSLLENKEAVLPAQDGLSEISFPNMSLQDHE 4891
Qy 1098 --LSDE-----TNI-----CNGKPVDTLTLRNGT 1120
Db 4892 LILSSQPHGWITHPNTPGKIYLYHYAEPPTGPFVEPPDLFFLKTTKSPVQ-----SPT 4946
Qy 1121 LVA-----FRGH-----YFWMLSPF 1135
Db 4947 LIAKSPREMVSRPEYKQALLPGHGGKNSRAPNMSLQDQLTSSQPHGW----- 5001
Qy 1136 SPSPARRITEVWGPSPIDTTFTRCNCEGKT-----FFKDSQ----- 1174
Db 5002 -----IPHPENT-----PGKIYLYHYAEPPTGPFVEPPDLFFLKTTKSKP 5040
Qy 1175 -YWRFTNDIKAGYKPIFKGFGGLTGQIVAAALSTAKYK----- 1212
Db 5041 VQW-----TSAEIAKLPEIVS-----QTKHAKESVLRAPVEOHKNSPPPNVS 5084
Qy 1213 -----NW-----PESVYFFKRGSGIQOYIY 1232
Db 5085 LQPLDQLSLFSQPHGWIPHPNTPGKIYLYHYAEPPTGPFVEPPDLFFLK-----TT 5136

QY	1233	KQBPVQKCP	-----GRRPA-----	1246
		: : : :	: :	
Db	5137	KSPVQGTPTYMTKSPQEIYVQSPQEGMEAGPPTKVEQDESPSPHASLQTLQELTSS	5196	
		: : : :	: :	
QY	1247	-----LNY-----	-----PVYGMTQVRRRPERA	1266
		: :	: : :	
Db	5197	QPHGWIPHPENTAGKIYLVHAEPTGPVFEPPDLFFLTKTKSPVQGTATQSTQ	5250	
		: : : :	: : : :	
QY	1267	IGFSQTHIRIQSPARLAYQDKVLHNEVKVSIILWRGLNPNVTSAI	-----SLFNIR	1319
		: : : :	: : : :	
Db	5251	-----SPKEI-----	-----ISPSTRYKESVLTAAVTQDESQSPNIS	5284
		: : : :	: : : :	
QY	1320	KPDGYDYAFKSDQYINIDVPSRTARAITRSGQTLKSVW	1359	
		: : : :	: : : :	
Db	5285	LOPDDOELTTLISOPHVLIIHSPSS	-----NANTGHTLGKIY	5320
		: : : :	: : : :	

RESULT 12

RESOBI
S49915

343313
extensin-like protein - maize

C: Species: Zea mays (maize)

C/species: Zea mays (maize)
C/Date: 05-Mar-1995 #sequence revision 12-May-1995 #text change 09-Jul-2004

C; Date: 05-Mar-1995
C; Accession: 848915

C;Accession: S49915
P:Publication: A T : Broadwater A H : Lowrey K : Redinger P A :

R;Rubinstein, A.L.; Broadwater, A.H.; Lowley,

submitted to the EMBL Data Library, June 1994

A;Description: Pex genes:

A;Reference number: S.

A;Accession: S49915

A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PDB:CAJ84230.1; PTD:96001

Query Match
12.3%: Score 898.5; DB 2; Length 1188;

Query Match	12.5%, score 330.3, 22
Best Local Similarity	20.5%: pred. No. 1.7e-11:

BEST LOCAL SIMILARITY	20.3%	PREC. NO. 117
Matches	278	Conservative
	91	Mismatches
	Indels	625
	Gaps	55

[illegible]

48 y -----QCKKYD 53
 | | :
 | | :

54 KCCPDYESFCAEVHNPT---SPSS-----KKAPPSGASQTIKSTTYESPKPPNKK 103

104 TTKVIESEETIEHSVSENQESSSSSSSSSTIWKSSKNSAANRELOKLVKVDNK 163

164 KNRTRKKKPTPKP-----PVDGAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAK 215
 440 ----VSPEPLPEPSPV-----PVDGAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAK 215
 440 ----VSPEPLPEPSPV-----PVDGAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAK 215

452 -----PAPAPMWTSPHPAD-----DYVPTTPVPGKSPPTATSPSPQVQPPA 495

496 ASTPPPSL-----503

```

504 --VKLSPQAPVGSPPPPVKTSTSPAPIGSPSPPPPVSVSPPPPVKSPPPPAVGSPPP 561

```

562 PEKSPPPAPVASPPPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVKSPPPPTVA 621

```

362 EPAPTTTKEPAPTTTTKEPAPTTTWSAPITP-KEFAPITPAKPAFAPITPKKPAFAPIT-
2Y
622 SPPPPAPVASSPPPMKSPPPPTTFVSSPPPPPEKSPPPPPAKSTPPPEEYPTPTTSVKSP 681
2b

```

DECSIII.T 13

RESULT
PAGE 66

B48666 cell proliferation antigen Ki-67, short form - human

cell proliferation antigen (man) C-reactive protein (man)

C;Species: Homo sapiens (Mall)
 a.date: 03 May-1994 #sequence revision 03-May-1994 #text change 15-Mar-2004

C;Date: 03-May-1994
C-Registration: B18666

C;Accession: B48666
p.Schluter C.: Duchrow M.: Wohlenberg C.: Becker. M.H.G.: Kev, G.: Fläd, H.D.: Ger

R;Schlueter, C.; Duchrow, M.; Wollenberg, C.; Becker, M.H.S.; Key, G.; Tag, A.S.;
T Cell Biol 123 512-522 1993

J. Cell Biol. 123, 513-522, 1993

A;Ti

INS. 01043425. PMID-8227122

A; Reference number:

A;Accession: B48666

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2897 <SCH>

C;Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
 F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 12.3%; Score 897.8; DB 2; Length 2897;
 Best Local Similarity 14.8%; Pred. No. 1.2e-10;
 Matches 464; Conservative 153; Mismatches 476; Indels 2052; Gaps 115;

```

QY 3 WKT-----LPIVLLLLSVFVIQVSSQELSCGRCFSEFREGRECDCA 47
Db 2 WPTRRLLVIKRGVDGPHFPLSLTCL-----FGRIECDIRI 39
QY 48 Q-----CKYDKCCDYESFCAEVH-----NPT-----70
Db 40 QLPVVSKQHCK-----IEIHQEAILLHNSFTNPTQVNGSVIDEPVRLKHGD 86
QY 71 -----SPPSSKKAPPSGASQT 87
Db 87 VITIIDRSFRYENSLONGRKSTFPPKIRIQEAPRRVSRSSFSDDPEKAQDSKAYSKI 146
QY 88 IKSTTKRSKP-----PPN-----KKKT-----KKVISEEITEHSYSEN 122
Db 147 TEGKVSNGPELFDENLPNPLKGEAPTCKKSLVMTHTPPVLKLIK-----EQQPSGK 201
QY 123 QESS-----SSSSSSSSST-----137
Db 202 QESGSEIHVEYKAQSLVISPPAPSPRKTPVASDQRRRSCKTAPASSKSQTVEVPKRGGER 261
QY 138 -----IWKIKSKNSAANRELOK-----KLVKDNK-----163
Db 262 VATCLOKRVISRSQHDILOMICSKRRSGASEANLIVAKGWADVVKLGAKQTQTKVIKHG 321
QY 164 ---KNTTKKKP-TPKPPV-----177
Db 322 PQRSMNKQRRRPATPKPKPVGEVHSQFSTGHANSPTIILIGKAHTEKVHVHPARVRLNNF 381
QY 178 -----VDEAG-----SGLDNGD-----189
Db 382 ISNQKMDFKEDLSGIAEMFKTPVKEQQLTSTCHIALNSENLLGKQFGQTDSEELLP 441
QY 190 -----FKVT-----TPD 196
Db 442 TSSEFGNVFFSQNAKQSDKCSAPPLRQCIRENGNVAKTPRNTYKMTSLETKTS 501
QY 197 TSTQHNKYST---SPKITAKPINRP-----221
Db 502 TETPSKTVSTVNRSGRSTEFNIOKLPIVESKSEETNTEIVEICILKRGQXATILQORREG 561
QY 222 -----SLP-----224
Db 562 EMKEIERPPFETYKENIELKENDEKMKAMKASRTWGQKCAPMSDLTDLKSLPDTELMKDTA 621
QY 225 -----PNSDTSKETSLTVNK-----ETTVETK-----246
Db 622 RGONLLQTDHAKAPKSEKGIKTMPCQSLOPEPINTPTHTKQOLKASLGKVGKBEELLA 691
QY 247 -----ETTNTKQSTDGKE-----TKTSAKETOSTE 273
Db 682 VGKFRITSGETTHTHREPAGDGKSIRTFKESPKQILDPAARVTGMKKWPTPKBEAQSL 741
QY 274 KTSAKDLATPSKVLAQTPKAET-----TTK-----299
Db 742 -----DLAGFKELFQTPGSPSESMDEKTKTKIACKSPPPESVDTPTSTKQWPKRSLRKAD 796
QY 300 -----GPALITPKPEP-----TPT-----TPK 315
Db 797 VBEFELALRLKLTSPSAGKAMLTTPKAGDEKODIKAFMGTPVQKLDLAGTLPGSKRQLQTEK 856
QY 316 EPAST-----TPKEP-----TPTTIKSAPT-----335
Db 857 EKAQALEDLAGFKELFQTPGHTHEELVAAGTKIKIPCDSPQSDPVDPTPTSTKQPKRSIRK 916
QY 336 -----TPKEPAP-----TTTAKSA 348

```

```

Db 917 ADVEGELLACRLNMPISAGKAMHTPK---PSVGEKDIIFVGTPTVQKLDLTENLTGSKRR 973
QY 349 PTTTPKEPAPT-----TTKEPAPTTTPKEPAPT-----383
Db 974 PQTPEEAQALEDLTGFKELFQTPGHTHEEVAAGTKTKWPCSESPESADT-----PTS 1027
QY 384 TKSAPTTTPKE-----PA 395
Db 1028 TRRQPKTPLEKRDYQKELSALKLQTSGTTHDKVPGGEDKSINAFRETAQOKLDPA 1087
QY 396 PTTTPKAPTTTPKEPA-----PTTPK-----EPTPT-TPK 424
Db 1088 SVTQSKHPT-KEKAQPLEDLAGFKELFQTPVCTDRPTTHEKTKTIACKSQPDVPTPT 1146
QY 425 EPAPTTKE-----PAPT-----436
Db 1147 SSKPQSKRSLRKVDVEEFALRKRTSPSAGKAMHTPKPVSAGEKNIAFMGTPTVQKLDLT 1206
QY 437 -----TPKEPA-----PTAPKK 448
Db 1207 ENLTGSKRRLQTPKEKAQALEDLAGFKELFQTRGHTBESMTNDKTAKVACKSQPDLDKN 1266
QY 449 PA-----PTTPKEPAPT-----460
Db 1267 PASSKRRLKTSLGKVGKVELLAVGKLTQTSGETTHTEPTGDKSKMAMFESPKQILD 1326
QY 461 -----TPKEPA-----PTTPKEPAPT-----481
Db 1327 SAASLTGSKRQLRTPKGSVPEDLAGFIELFQTPSHTK-SMTNEKTKTVSVASQPD 1385
QY 482 --APTTPKSAP-----490
Db 1386 VDTPTSSKQPKRSLRKADTEEBFLAFRKQTPSAGKAMHTPKPAVGEKDINTFLGTPVQ 1445
QY 491 -----TTTKEPA-----PT-----507
Db 1446 KLDQPGNLPGSNRLQTRKEKAQALELTGFRFELFQTPCTDNPTADEKTTKKILCKSPQS 1505
QY 508 PKEPSPPTTKE-----PA-----520
Db 1506 DPADTPTNTKQPKRSLKKADVEEFALFRKLTPSAGKAMHTPKAAVGEKDINTFVGTP 1565
QY 521 -----PTTPKEPAPT-----TP-----KKP 535
Db 1566 VEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKELFQTEGHTESMTDDKI TEVSCKSP 1625
QY 536 APTTPKEPA-----544
Db 1626 QPDVPVKTPTSSKQRLKISLGKVGKVEEVLPGKLTQTSKTTQTHRETAGDGKSIKAFKE 1685
QY 545 -----PTTPKEPA-----PTTTKPAAPTAPK 565
Db 1686 SAKQMLDPANYGTGMERWPRTPKEEAQSLDLAGFKELFQTPDHTESITDDKTTKIACK 1745
QY 566 EPAP---TTTPKET--APTTP-----KLTPTT-----587
Db 1746 SPPPSMDPTSTRRRPKTPLGKRDI VEELSALKQLTQTHTDKVPGEDKGINVPRETA 1805
QY 588 PEKLAB---TTTEKPAPTTP-----BELA-----PTTSEEPPTTPPEEP 623
Db 1806 KQKLDPAASVTGSKRQPTPKQAQPLEDLAGFKELFQTPVCTDRPTTHEKTKTIACKRSP 1865
QY 624 AP-----TTPKAAA-----632
Db 1866 QPDVPVGTPTIFKPQSKRSLRKADVEEESLALRKRTSPSVGKAMDTPKPAGDEKMKAFMG 1925
QY 633 -----PNTPKPEPA-----PTTPKEPAPTTPKE 654
Db 1926 TPVQKLDLPGNLPGSKRWPTPKQAQALEDLAGFKELFQTPGTDKPTTDEKTTKIACKS 1985
QY 655 PAP---TTTPKET-----APTTPK-----669
Db 1986 PQDPDPVDTPTASTKQPKRNLKADVEEFALRLKRTSPSAGKAMDTPKPAVSEKNTFTV 2045

```

670 QY -----GTAPTTLKBPAPTPPKKAP-----KELAPTTTKEPTSTSDK----- 707
Db : : : : :
2046 ETPVQKLDLGNLPGSKRQ--POTPKKAEALDLVGFELFQTPGHTBESMTDDXITEV 2103
QY -----PAP-----TPPKGTAPT----- 719
Db : : : : :
2104 SCKSPQSFKTSRSKQRLKIPLVKDMKEBLAVSKLRTSTGETTQHTFTGDSKSI 2163
QY -----TPK-----EPAPT-----TPKEPA----- 733
Db : : : : :
2164 KAFKESPKQLDPAASVTGSRRLRTKKEKARALEDLVDFKELFSAPGHTESMTIDKNT 2223
QY -----PTPKGTATPTTLKBPAPTPPKKAPKELAP-----TTTKGPTSTTSKPA----- 778
Db : : : : :
2224 KIPCKSPPLDTATSTKRCP--KTRPRKEVKELSAVERLTQTSGQSTHKEPAGSDE 2282
QY -----PTTPKE-----TAPTT 789
Db : : : : :
2283 GIKVLKORAKKKNVEEESRRRPRAPKKAQPLEDLAGFTLSETSHTQESLTAGKA 2342
QY PKBPAPTPPKKAPTPPETPTTTSVSTPTTTKBTTHKSPDESPELSABTPP----- 845
Db : : : : :
2343 TKIPCESPPLVVDVTTASTKHLRLTRVQKVQKEEPSAV--KFTQTSGETTADAKEPAGED 2401
QY -----KALENSKEPGVPTT-----KTPAA--TKPEMTTAK 875
Db : : : : :
2402 KGKALKESAKQTPAPAAASVTGSRRLRPRAPRESAQALEDLAGFKDPAAGHTESMT--D 2458
QY DKTERDLRTPE--TTTAAPKMTKETAATTTKTE--SKITATTTQVTSITTQ-- 925
Db : : : : :
2459 DKTIKIPCKSPELEDATSKRPRPRTRAKQVEKVELLAVGKLTQTSGETTHDKPEVG 2518
QY -----DTPPKKLTTLTKTLAPKVTTKTITTEIMNKPEETAKPKORA----- 970
Db : : : : :
2519 EGKTKAFK-----QPAKRVDAEDVIGSRROPRAPKKAQPLEDLASQEL 2565
QY -----TNSKATTPKQKPTK-----APK----- 988
Db : : : : :
2566 SOTPGHTELANGAADSFSAKQTPDGSGLKLSRVLRAKVPVGVVDVSTDRPVKSQ 2625
QY -----KPTSTKPKTPRVRKPTTTPRKMSTMP-----ELNP 1023
Db : : : : :
2626 SKSNTSLPPLPFKRGKGDSVGTGKRLCM-----PAPEIIEELPASKKQVAP 2676
QY -----IARAMLQTTTR-----PNQTPNSKLVEVP-----KSEDAGGAGETPHMLLR 1069
Db : : : : :
2677 RARGKSSBPVIMKRSRLTSAKRIEPAEBELNSDMKTNKEEHLQDS----- 2723
QY PHVFMPEVTPDMVLPVFNQGIILNPMLSDETNI CNKGPVDGLTLNGLVAFRGHYF 1129
Db : : : : :
2724 -----VPE-----NKGISLRSRQDKT----- 2740
QY WMLSPFSPSPARRITEVW-----GIPSPIDTVFTRCNCEGT 1167
Db : : : : :
2741 -----EAEQILTEVFLAERIEINREKPKMTSPENDIQNP----- 2777
QY FFFKDSQYWRFTNDIKDAGYKPIPKGPGGLTGQIIVAALSTAKYNWPESVFFKRGSI 1227
Db : : : : :
2778 -----DDGARKPI----- 2785
QY QQYIYKQEVQKCPGRRPALNVPVVGEMTQ-----VRRRRFERAIGPSQ 1271
Db : : : : :
2786 -----PRDKVTENKRLSARQNESQPKVAEESGGQSAKVMQNKQKGGAGNSD 2837
QY THTR-----IQYSPARLAYQDGLVHNEVKVLSILMRGLPNVVTSAIS--LPNIRKPDGYDY 1326
Db : : : : :
2838 SMCLRSRKTQSQAASTLESKSVQR-----VTRSVKRCAPENPKKAED----- 2879
QY VAFSKDQYNIQVPSRTARAITRS 1351
Db : : : : :
2880 -----NVCV-----KKITRS 2890

RESULT 14

225697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: the sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-1229 <Full>
A:Cross-references: UNIPROT:Q94195; EMBL:U67956; PIDN:AA07691.1; GSPDB:GNO0028; CESP:F16F9
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.2%; Score 890.1; DB 2; Length 1229;
Best Local Similarity 23.8%; Pred. No. 2.8e-11;
Matches 372; Conservative 116; Mismatches 395; Indels 679; Gaps 76;

QY 2 AWKTLPIYLLLLSVFVIQQVSSQELSCKG-RCFESPGRGECDDCAQCKKYDKCCPDYE 60
Db 4 AW-VVSPAFILGNVQSSLLSKTINSQSRDFKIVKHKKNTCTCSC---KCVPD-- 56
QY 61 SFCAEVNHTSPSPSKKAPPPSGASQTIKTTKRSPKPNKKTKKVIKIEEITEHSVS 120
Db 57 -----APSN-----PFDVSTTSSIN-----N 73
QY 121 ENOESSSSSSSSSSSTIW-----KIKSSKNSAANRELOKLVKDNKNKRTKK 169
Db 74 DNDVIGSGDSNPTGSS-WFQIEATVGGQTVKSEHNIDSSVEVEKKV----- 120
QY 170 KPTPKPVVDEAGSLDNGDFKVTTPDTSTTQ-----HNKVTSPKI----- 211
Db 121 -----TTSDASITNAFTCKDSTTPEIITGIVVINSKESVT 158
QY 212 -----TTAKPINRPSLPNSDTSKETS LVNKEETVETKETTNTKQTSDDGKEKT 263
Db 159 DMSTTSTSTLSP-----TTLELTSPELTVSDSTST-EQTSPTDNTTPEI 202
QY 264 TSAKETOSIEKTSADLAPT SKVLAKTPPRAETTTKGPAITTPKEPTTT--PKEPAST 320
Db 203 ASPMET-----NTTTE--ATTTSVEPSVSTLASEDEITVT 235
QY 321 TPKEPTPTTIKSAPTTPKEP-----APTITKSAPTTPKEPAPTTTPKEPAPTTTPK 377
Db 236 AIAESTTVIAEVSTTTEPTTAEITTKGTT---KAPATTEPTTTEE--VTTTE 289
QY 378 EPATTTKSAPTTPKEPAPTPP-----KKPAPTTTPKEPAPTTTPKEP--TPTTPKEPAPTKE 432
Db 290 AETSTTTSSETSTSK-----PTTPLIDNKIAGPATGK---PETTHFPVGTGTP--NFD 337
QY 433 PAPTPP-----KEPAPTAAPKAPTTTPKEPAPTTTPKEPAPTTTKE-----PSPTTPKEPAP 483
Db 338 TATETPFVAKSEDKMTLSKTAATETQOTTEVT-DGPEKETIKVNSIETPITTVPLVET 396
QY 484 TTTKSAP-----TTTKEPAPT-----TTKSAPTTPEPSPTTKE-----PAPT----- 522
Db 397 STTSTASKESDGEHTLLKLVTTADSDSTESATTVKPENEETTTKSHVVPKTKGTVKV 456
QY 523 TPK-----BPAPTTPKKAP----- 537
Db 457 TPKLELSEDEPTEIT-KAPHECKLLEKTYHVLSDNFARVSEAKENDDYHLDYNNHRE 515
QY 538 -----TTTKEPA-----PTTPKEPAPTTTTPKAPTAPEKA--PTTPKE 574
Db 516 AKBPITTEESSTTEVTTEEPANTGNPTTEN---PTTTEQPTSTABSTTALPFTTEQ 572

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 109.728 Seconds
(without alignment)
4449.477 Million cell updates/sec

Title: SEQ1-G

Perfect score: 7323

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLKQVWYVNC 1361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7308.7	99.8	1404	2	AAR26049 MSF precu
2	7308.7	99.8	1404	4	AAB29773 Human meg
3	7308.7	99.8	1404	4	AAB60568 Human meg
4	7308.7	99.8	1404	8	ADM98014 Human meg
5	7305.7	99.8	1415	4	AAU32262
6	7293.7	99.6	1404	7	ADK55839
7	7043.9	96.2	1320	7	ADK65819
8	7043.9	96.2	1311	8	ADK67912
9	6778.9	92.6	1270	8	ADK67911
10	6732.7	91.9	1299	4	AAW24322
11	3552	48.5	902	4	AAB29778
12	2929	40.0	551	4	ABUS3253
13	2920	39.9	546	4	ABUS3252
14	2850.9	38.9	538	5	AAO18834
15	2757	37.6	513	4	ABUS3254
16	2232.9	30.5	452	2	AAR80041
17	2004	27.4	472	4	AAB60569
18	1566.9	21.4	292	5	AAU11261
19	1413.6	19.3	5179	4	AAW24516
20	1413.6	19.3	5179	6	ABP55365
21	1413.6	19.3	5179	6	ABO07258
22	1413.6	19.3	5179	7	ADDA48091
23	1413.6	19.3	5179	7	ADDA4998
24	1189.7	16.2	8951	6	ABU08487
25	1179	16.1	1664	2	AAW43106

26 1139.6 15.6 5703 8 ADL23265 Human MUC
27 1117 15.3 1795 4 ABB69806 Drosophil
28 1112.5 15.2 4315 5 ABP43908 MUC5B par
29 1100.5 15.0 717 4 ABUS3144 Human tes
30 1097 15.0 214 4 ABUS3255 Human tes
31 1092.6 14.9 763 3 AAG38942 Arabidops
32 1080.9 14.8 2284 4 ABB71434 Drosophil
33 1048.5 14.3 1049 4 ABB61364 Drosophil
34 1026.2 14.0 4498 4 ABB58595 Drosophil
35 1010 13.8 188 5 AAO18833 5' cartil
36 1009.6 13.8 2768 4 ABB68397 Drosophil
37 995.4 13.6 2112 4 ABB60403 Drosophil
38 992.7 13.6 770 4 ABUS3141 Human tes
39 992.5 13.6 1460 7 ADN39110 Cancer/an
40 992.2 13.5 1538 7 ADI21202 Novel hum
41 987.6 13.5 778 4 ABUS3143 Human tes
42 978.6 13.4 1370 7 ADI21666 Novel hum
43 974.4 13.3 692 4 ABUS3155 Human tes
44 964.1 13.2 717 4 ABUS3150 Human tes
45 964.1 13.2 717 4 ABUS3149 Human tes

ALIGNMENTS

RESULT 1
AAR26049
ID AAR26049 standard; protein; 1404 AA.
XX AC AAR26049;
XX DT 25-MAR-2003 (revised)
XX DT 02-FEB-1993 (first entry)
XX DE MSF precursor.
XX KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
XX KW stability; proteolytic cleavage; adhesion; alternative splicing.
XX OS Synthetic.
XX PH Key
XX PH Region Location/Qualifiers
FT Region 1..26
FT Region /label= Exon_I
FT Region 26..67
FT Region /label= Exon_II
FT Region 67..107
FT Region /label= Exon_III
FT Region 107..157
FT Region /label= Exon_IV
FT Region 157..200
FT Region /label= Exon_V
FT Region 200..1141
FT Region /label= Exon_VI
FT Region 1166..1212
FT Region /label= Exon_VII
FT Region 1213..1266
FT Region /label= Exon_IX
FT Region 1266..1331
FT Region /label= Exon_X
FT Region 1331..1373
FT Region /label= Exon_XI
FT Region 1373..1404
FT Region /label= Exon_XII
FT Region 1411..1166
FT Region /label= Exon_VII
WO9213075-A1.
06-AUG-1992.
17-JAN-1992; 92WO-US000433.

Db 1201 RNCBEGKTFKDSQWRFNDIKDAGYKPKIPKGFGLTGQIVAAALSTAKYKNWESVY 1260

QY 1218 FFKRGGSIQYIYKQBPVQKCPGRPALNYPVYGEMLTQVRRRFRFAIGFSQTHIRIQY 1277

Db 1261 FFKRGGSIQYIYKQBPVQKCPGRPALNYPVYGEMLTQVRRRFRFAIGFSQTHIRIQY 1320

QY 1278 SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYNNIDV 1337

Db 1321 SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYNNIDV 1380

QY 1338 PSRTARAITRRSGTILSKWYNCP 1361

Db 1381 PSRTARAITRRSGTILSKWYNCP 1404

RESULT 3

AAB60568

ID AAB60568 standard; protein; 1404 AA.

XX AC AAB60568;

XX DT 27-APR-2001 (first entry)

XX DE Human megakaryocyte stimulating factor (MSF, CACP).

XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis; MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic; antiarthritic.

XX OS Homo sapiens.

XX PN WO200107068-A1.

XX PD 01-FEB-2001.

XX PF 21-JUL-2000; 2000WO-US020002.

XX PR 23-JUL-1999; 99US-0145328P.

XX PR 19-JUL-2000; 2000US-00145328.

XX PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX PI Warman ML;

XX PS WPI; 2001-182721/18.

XX PT New composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis protein in combination with an anesthetic, useful for treating osteoarthritis, or as lubricants of tissue and joints.

XX PS Example 1; Page; 34pp; English.

XX CC The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: this sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)

SQ	Sequence 1404 AA;	Query Match 99.8%; Score 7308.7; DB 4; Length 1404;
	Best Local Similarity 96.9%; Pred. No. 1.8e-176;	
	Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;	
QY	1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCNDYNCQHYMECCPDF 60	
Db	1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCNDYNCQHYMECCPDF 60	
QY	61 KRYCTAELSCKRCFESFERGRCDDAOCKKYDKCCPDYESFCAEVHNTSPSSKKAP 120	
Db	61 KRYCTAELSCKRCFESFERGRCDDAOCKKYDKCCPDYESFCAEVHNTSPSSKKAP 120	
QY	121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE----- 156	
Db	121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE----- 180	
QY	157 -----VKDNKQKRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST 197	
Db	181 KIKSSKNSAANRELQKKLVKDKNKQKRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240	
QY	198 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVWKETTVEKTTTNKQTSDDG 257	
Db	241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVWKETTVEKTTTNKQTSDDG 300	
QY	258 KEKTTSAKETQSIEKTSAKDLAPTSLVLAAPTPKAETTTKGPALTTTKEPTPTTKKPAS 317	
Db	301 KEKTTSAKETQSIEKTSAKDLAPTSLVLAAPTPKAETTTKGPALTTTKEPTPTTKKPAS 360	
QY	318 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 377	
Db	361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420	
QY	378 APTTKSAPTTPKEPAPTTTKKAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTK 437	
Db	421 APTTKSAPTTPKEPAPTTTKKAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTK 480	
QY	438 EPAPTAPKKAPATTPKEPAPTTTPKEPAPTTTKKPSPTTPKEPAPTTTKSAPTTTKEPAPT 497	
Db	481 EPAPTAPKKAPATTPKEPAPTTTPKEPAPTTTKKPSPTTPKEPAPTTTKSAPTTTKEPAPT 540	
QY	498 TTKSAPTTTPKEPAPTTTKEPAPTTTPKKAPATTPKAPATTPTPKEPAPTTTPKEPAPT 557	
Db	541 TTKSAPTTTPKEPAPTTTKEPAPTTTPKKAPATTPKAPATTPTPKEPAPTTTPKEPAPT 600	
QY	558 APTAPKGPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEKAPATTPPEELAPTTPEPTPTT 617	
Db	601 APTAPKGPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEKAPATTPPEELAPTTPEPTPTT 660	
QY	618 PEEEPATTPKAAAPNTPEKAPATTPKAPATTPKAPATTPKETAPTTPKGTAPTTLKEP 677	
Db	661 PEEEPATTPKAAAPNTPEKAPATTPKAPATTPKAPATTPKETAPTTPKGTAPTTLKEP 720	
QY	678 APTTPKKEPAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTPKAPATTPKAPATTPKGT 737	
Db	721 APTTPKKEPAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTPKAPATTPKAPATTPKGT 780	
QY	738 TAPTTLKGPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPKETAPTTPKAPATTPK 797	
Db	781 TAPTTLKGPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPKETAPTTPKAPATTPK 840	
QY	798 KPAPTPTPTPTTSEVSTPTTKEPTTIHKSPDSTPELSAEPPTPKALENSKPEFGVPT 857	
Db	841 KPAPTPTPTPTTSEVSTPTTKEPTTIHKSPDSTPELSAEPPTPKALENSKPEFGVPT 900	
QY	858 TKTPAATKPEMTTAKDKMTTBRDURTTTETTTAAPKMTKETATTTTEKTESKITATTQV 917	
Db	901 TKTPAATKPEMTTAKDKMTTBRDURTTTETTTAAPKMTKETATTTTEKTESKITATTQV 960	
QY	918 TSTTTQDTPPKITTLKUTTLAPKVTTKKTTTTEINMKPBETAKPKDRATNSKATTPK 977	
Db	961 TSTTTQDTPPKITTLKUTTLAPKVTTKKTTTTEINMKPBETAKPKDRATNSKATTPK 1020	

QY 978 PQPTKAPKPTSTKKPTMPVRKPTTPTRKMTSTMPELNPTSRIABAMLOTTTPN 1037
 DB 1021 PQPTKAPKPTSTKKPTMPVRKPTTPTRKMTSTMPELNPTSRIABAMLOTTTPN 1080
 QY 1038 QTPNSKLVEVNPKEADAGAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS 1097
 DB 1081 QTPNSKLVEVNPKEADAGAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS 1140
 QY 1098 DETNLCNGKPYDGLTTLNGLTVAPRGHYFWMLSPFPSPARRITEVWGIPIIDTVFT 1157
 DB 1141 DETNLCNGKPYDGLTTLNGLTVAPRGHYFWMLSPFPSPARRITEVWGIPIIDTVFT 1200
 QY 1158 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQIIVAAALSTAKYKNWPESVY 1217
 DB 1201 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQIIVAAALSTAKYKNWPESVY 1260
 QY 1218 FFKGGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1277
 DB 1261 FFKGGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1320
 QY 1278 SPARLAYODKGLVHNEVKVSLWRGLPNVVTSAISLENIKRPDGDYDYAFKSDQYINIDV 1337
 DB 1321 SPARLAYODKGLVHNEVKVSLWRGLPNVVTSAISLENIKRPDGDYDYAFKSDQYINIDV 1380
 QY 1338 PSRTARAITTRSGQTLKSVWYNCP 1361
 DB 1381 PSRTARAITTRSGQTLKSVWYNCP 1404

RESULT 4

ADM98014
 ID ADM98014 standard; protein; 1404 AA.
 AC ADM98014;
 DT 01-JUL-2004 (first entry)
 XX Human megakaryocyte stimulating factor (MSF).
 XX lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
 KW CAP; camptodactyl-arthritis pericarditis; osteoarthritis; human;
 KW megakaryocyte stimulating factor; MSF.
 XX Homo sapiens.
 XX US2004072741-A1.
 PD 15-APR-2004.
 PF 02-JUL-2001; 2001US-00897188.
 XX 23-APR-1999; 99US-00298970.
 PR 24-APR-2000; 2000US-00556246.
 XX (JAYG/) JAY G D.
 PA Jay GD;
 PI WPI; 2004-373948/35.
 XX N-PSDB; ADM98015.
 DR New tribonectin polypeptides and polynucleotides for lubricating joints
 XX or other tissues to prevent or treat Camptodactyl-arthropathy -
 PT pericarditis syndrome or osteoarthritis.
 PT Claim 1; SEQ ID NO 1; 34pp; English.
 PS The invention relates to a lubricating polypeptide and at least one O-
 CC linked oligosaccharide. The composition and methods are useful for
 CC lubricating joints or other tissues to prevent or treat camptodactyl-
 CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
 CC sequence represents the amino acid sequence of the human megakaryocyte

CC stimulating factor (MSF).
 XX SQ Sequence 1404 AA;
 Query Match
 Best Local Similarity 99.8%; Score 7308.7; DB 8; Length 1404;
 Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
 QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
 QY 61 KRVTAEELSCRCGFESFERGECDDAQCKYDKCCPDYSEFCAEVNHNPTSPSSKKAP 120
 DB 61 KRVTAEELSCRCGFESFERGECDDAQCKYDKCCPDYSEFCAEVNHNPTSPSSKKAP 120
 QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEETI----- 156
 DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEETI----- 180
 QY 157 -----VKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
 DB 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
 QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTIVETKETTNNKQTSDDG 257
 DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTIVETKETTNNKQTSDDG 300
 QY 258 KEKTTSAKETQSIKETSADOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
 DB 301 KEKTTSAKETQSIKETSADOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
 QY 318 TTPKEPTTTIKSAPTTKPEAPTTTSAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEP 377
 DB 361 TTPKEPTTTIKSAPTTKPEAPTTTTSAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEP 420
 QY 378 APITTKSAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 437
 DB 421 APITTKSAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480
 QY 438 EPAPTAPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 497
 DB 481 EPAPTAPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540
 QY 498 TTKSAPTTKPEPSTTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 557
 DB 541 TTKSAPTTKPEPSTTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600
 QY 558 APAPTAPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 617
 DB 601 APAPTAPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660
 QY 618 PEEPAPTTPKAAAPNTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 677
 DB 661 PEEPAPTTPKAAAPNTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720
 QY 678 APPTPKKAPKELAPTTTKEPSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 737
 DB 721 APPTPKKAPKELAPTTTKEPSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780
 QY 738 TAPTTTKEPAPTTTKEPAPKELAPTTTKEPSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEP 797
 DB 781 TAPTTTKEPAPTTTKEPAPKELAPTTTKEPSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEP 840
 QY 798 KPAPTTPPTPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAPPTPKALENSKEPQVPT 857
 DB 841 KPAPTTPPTPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAPPTPKALENSKEPQVPT 900
 QY 858 TKTPAAKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETATTTTEKTTESKITATTTQV 917
 DB 901 TKTPAAKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETATTTTEKTTESKITATTTQV 960
 QY 918 TSTTTTQDTTFFKLTITLTKTTLAPKVTITTKTITTTTINMKNPDEATKPKORATNSKATTPK 977

Db 961 TSTTTQDTDFKLTLLKTTTLAPKVTITTKITITTEINNKPEETAKKORATNSKATPK 1020
QY 978 PQKPTKAPKPTSTKPKTMRVRKPKTTPTRKWTSTMPENLPTSRABAMLQTTTRPN 1037
Db 1021 PQKPTKAPKPTSTKPKTMRVRKPKTTPTRKWTSTMPENLPTSRABAMLQTTTRPN 1080
QY 1038 QTNPKLVEVNPKSSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1097
Db 1081 QTNPKLVEVNPKSSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1140
QY 1098 DETNIONGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITTEWGLPSPIDTVFT 1157
Db 1141 DETNIONGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITTEWGLPSPIDTVFT 1200
QY 1158 RCNCEGKTFKFDKSYWRFTNDIKDAGYKPIFKFGGLTGQIIVAALSTAKYKNWPSVY 1217
Db 1201 RCNCEGKTFKFDKSYWRFTNDIKDAGYKPIFKFGGLTGQIIVAALSTAKYKNWPSVY 1260
QY 1218 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRFRERAIQSGQTHIRIQY 1277
Db 1261 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRFRERAIQSGQTHIRIQY 1320
QY 1278 SPARLAYQDKGVLHNEVKVSLWRGLPNVWTSALSLEPNIRKPDGYDYAFSKOQYXNIDV 1337
Db 1321 SPARLAYQDKGVLHNEVKVSLWRGLPNVWTSALSLEPNIRKPDGYDYAFSKOQYXNIDV 1380
QY 1338 PSRTARAITRSGQTLKSVWYNCP 1361
Db 1381 PSRTARAITRSGQTLKSVWYNCP 1404
RESULT 5
AAU32262
ID AAU32262 standard; protein; 1415 AA.
AC AAU32262;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2753.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-611725/70.
DR
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
PT
XX
PS Claim 20; Page 573; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 1415 AA;
Query Match 99.8%; Score 7305.7; DB 4; Length 1415;
Best Local Similarity 96.9%; Pred. No. 2.1e-176;
Matches 1360; Conservative 1; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 12 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 71
QY 61 KRVCCTAELSCKGRGCFESFERGECDCDAQCKKDYKCCPDYESFCAEVHNPTSPSSSKAP 120
Db 72 KRVCCTAELSCKGRGCFESFERGECDCDAQCKKDYKCCPDYESFCAEVHNPTSPSSSKAP 131
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIREEITE----- 156
Db 132 PPSGASQTIKSTTKRSPKPNKKTKKVIREEITEEHSVSENESSSSSSSSSSSTIW 191
QY 157 -----VKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
Db 192 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 251
QY 198 TQHNKVTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEVTVETKTTTTNQTSDG 257
Db 252 TQHNKVTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEVTVETKTTTTNQTSDG 311
QY 258 KEKTTSAKETQSIKTSKADLAPTSKVLAKTPKAEITTKGPAITTPKEPTTPPKEPAS 317
Db 312 KEKTTSAKETQSIKTSKADLAPTSKVLAKTPKAEITTKGPAITTPKEPTTPPKEPAS 371
QY 318 TTPKEPTPTTKSAPPTPKEPATTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 377
Db 372 TTPKEPTPTTKSAPTTPKEPATTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 431
QY 378 APTTTKSAPTTPKEPATTTPKKPAITTPKPEAPTTTPKEPTPTTKPEAPTTKEAPTTPK 437
Db 432 APTTTKSAPTTPKEPATTTPKKPAITTPKPEAPTTTPKEPTPTTKPEAPTTKEAPTTPK 491
QY 438 EPAPTAPKKAPATTPKPEAPTTPKPEAPTTTKKPSPTTKPEAPTTTKSAPTTTKPEAPT 497
Db 492 EPAPTAPKKAPATTPKPEAPTTPKPEAPTTTKKPSPTTKPEAPTTTKSAPTTTKPEAPT 551
QY 498 TTKSAPTTPKPSPTTKPEAPTTPKKPAITTPKPEAPTTTPKEAPTTTKPEAPTTTKKP 557
Db 552 TTKSAPTTPKPSPTTKPEAPTTPKKPAITTPKPEAPTTTPKEAPTTTKPEAPTTTKKP 611
QY 558 APTAPKEAPATTPKETAATTPKKUTPTTPPEKLAFTTPPKPAITTPPEELAPTTPEEPTTT 617
Db 612 APTAPKEAPATTPKETAATTPKKUTPTTPPEKLAFTTPPEKLAFTTPPEELAPTTPEEPTTT 671
QY 618 PEEPAATTPKAAANPTPKPEAPTTPKPEAPTTTKETAATTPKETAATTPKGTAPTTLKPE 677
Db 672 PEEPAATTPKAAANPTPKPEAPTTPKPEAPTTTKETAATTPKETAATTPKGTAPTTLKPE 731
QY 678 APTTPKPAKELAPTTTKETSTTSKPAITTPKGTAPTTPKPEAPTTTPKEAPTTPKG 737
Db 732 APTTPKPAKELAPTTTKETSTTSKPAITTPKGTAPTTPKPEAPTTTPKEAPTTPKG 791
QY 738 TAPTTLKPEAPTTPKKPAKELAPTTTKGPTSTTSKPAITTPKETAATTPKETAATTPK 797

Db 792 TAPTTLKEPAPPTPKKPAKELAPTTTKGPTSTTSDKDAPTTTKETAPTTTKEPAPPTPK 851
 Qy 798 KPAPTTPTPTTSEVSTPTTTKEPTTIHKSPTSTPSELSAEPPTPKALENSPKPGVPT 857
 Db 852 KPAPTTPTPTTSEVSTPTTTKEPTTIHKSPTSTPSELSAEPPTPKALENSPKPGVPT 911
 Qy 858 TKTPAATKPEMTTAKDKITTERDLRTTPTTTAAPKMTKETATTTTEKTESKITATTQV 917
 Db 912 TKTPAATKPEMTTAKDKITTERDLRTTPTTTAAPKMTKETATTTTEKTESKITATTQV 971
 Qy 918 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPK 977
 Db 972 TSTTTQDTPPKITTLKTTILAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPK 1031
 Qy 978 POKTPKAPKPTSTKKPKTMRVRPKTTPRKMTSTMPELNPTSRIAEAMLQTTTPN 1037
 Db 1032 POKTPKAPKPTSTKKPKTMRVRPKTTPRKMTSTMPELNPTSRIAEAMLQTTTPN 1091
 Qy 1038 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVYLPRVFNQGIINPMLS 1097
 Db 1092 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVYLPRVFNQGIINPMLS 1151
 Qy 1098 DETNICNGKPDVGLTTLNGLVAPRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT 1157
 Db 1152 DETNICNGKPDVGLTTLNGLVAPRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT 1211
 Qy 1158 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIEKFGGLTGOIIVAALSTAKYKNWPESVY 1217
 Db 1212 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIEKFGGLTGOIIVAALSTAKYKNWPESVY 1271
 Qy 1218 FFKRGSSIQQYIYKQEPVQKCGRRPALNYPVYGEMTQVRRRRPERALGPSQTHIRIQY 1277
 Db 1272 FFKRGSSIQQYIYKQEPVQKCGRRPALNYPVYGEMTQVRRRRPERALGPSQTHIRIQY 1331
 Qy 1278 SPARLAYQDKGVLEHNEVKVSIILWRGLPNVVTSAISLNPDKDGYDYVAFSKQYVYNDV 1337
 Db 1332 SPARLAYQDKGVLEHNEVKVSIILWRGLPNVVTSAISLNPDKDGYDYVAFSKQYVYNDV 1391
 Qy 1338 PSRTARAITRRSGQTLKSWVWNCV 1361
 Db 1392 PSRTARAITRRSGQTLKSWVWNCV 1415

RESULT 6

ADK65839
 ID ADK65839 standard; protein; 1404 AA.
 XX
 AC ADK65839;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Angiogenesis-differentially expressed protein #53.
 DE
 KW cyostatic; cardiac; vasotropic; antiarteriosclerotic;
 KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
 KW gene expression; cancer; coronary artery disease; myocardial ischemia;
 KW coronary arteriosclerosis; forensic medicine.
 XX
 OS Homo sapiens.
 XX

FN WO2003066831-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003848.

XX 07-FEB-2002; 2002US-00067482.

PR 10-JUN-2002; 2002US-00164595.

PR 16-AUG-2002; 2002US-0403649P.

PR 03-JAN-2003; 2003US-0437746P.

XX (ORIG-) ORIGENE TECHNOLOGIES INC.

XX

PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;

XX WPI; 2003-731502/69.

XX
 PT Determining the angiogenic index of a tissue or cell sample using
 PT expression levels of differentially expressed genes, useful for
 PT diagnosing or treating cancer, coronary artery disease, myocardial
 PT ischemia and/or arteriosclerosis.
 XX

PS Disclosure; SEQ ID NO 78; 296pp; English.

XX
 CC The invention relates to a method of determining the angiogenic index of
 CC a tissue or cell sample comprising assessing, in a sample, the expression
 CC levels of one or more differentially-expressed gene from any of 34 DNA
 CC sequences, given in the specification, where the levels are indicative of
 CC the angiogenic index. The methods and compositions of the present
 CC invention are useful for diagnosing, preventing and/or treating cancer,
 CC coronary artery disease, myocardial ischemia or coronary
 CC arteriosclerosis. They can also be used in research, drug discovery and
 CC forensic medicine involving angiogenesis. This sequence corresponds to
 CC one of the differentially expressed proteins of the invention.
 XX

SQ Sequence 1404 AA;

Query Match 99.6%; Score 7293.7; DB 7; Length 1404;
 Best Local Similarity 96.7%; Pred. No. 4.2e-176;
 Matches 1358; Conservative 0; Mismatches 3; Indels 43; Gaps 1;

Qy 1 MAWTLPIYLLILLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCVNCOHMECCPDF 60
 Db 1 MAWTLPIYLLILLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCVNCOHMECCPDF 60
 Qy 61 KRVTAEALSCKRCFESFERGECDCDAQCCKYDKCCPDYESFCAEHNPTSPSSKKAP 120
 Db 61 KRVTAEALSCKRCFESFERGECDCDAQCCKYDKCCPDYESFCAEHNPTSPSSKKAP 120
 Qy 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIREEITE----- 156
 Db 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIREEITE----- 180
 Qy 157 -----VKDNKNRTKKKPTPKPPVDEAGSLGNDGPKVTPDTST 197
 Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVDEAGSLGNDGPKVTPDTST 240
 Qy 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTDG 257
 Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTDG 300
 Qy 258 KEKTTSAKETQSIEKTSADLAPTSKVLAKETPKAETTTKGPALTTPKEPTTPPKEPAS 317
 Db 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKETPKAETTTKGPALTTPKEPTTPPKEPAS 360
 Qy 318 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 377
 Db 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 420
 Qy 378 APTTTKSAPTTTPKEPAPTTTPKKDAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPK 437
 Db 421 APTTTKSAPTTTPKEPAPTTTPKKDAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPK 480
 Qy 438 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTKSPSTTPKEPAPTTTKSAPTTTTKPAPT 497
 Db 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTKSPSTTPKEPAPTTTKSAPTTTTKPAPT 540
 Qy 498 TTKSAPTTPKEPSPTTTTKPAPTTTPKEPAPTTTPKKDAPTTTPKEPAPTTTPKEPAPTTTKKP 557
 Db 541 TTKSAPTTPKEPSPTTTTKPAPTTTPKEPAPTTTPKKDAPTTTPKEPAPTTTPKEPAPTTTKKP 600
 Qy 558 APTAPKEPAPTTTPKETAPTTTPKKLTPTTPEKLAAPTTPKEPAPTTPEELAPTTPEEPTPTT 617
 Db 601 APTTPKEPAPTTTPKETAPTTTPKKLTPTTPEKLAAPTTPKEPAPTTPEELAPTTPEEPTPTT 660
 Qy 618 PEEEPAPTTPKAAAPNTPKSAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGATPTTLKEP 677

QY 181 SGLDNGDFKVTTPDTSSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKETT 240
Db 131 SGLDNGDFKVTTPDTSSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKETT 190
QY 241 VETKETTTTINKQTSSTGKEKTSIAKETQSIKTSIAKOLAPTSKVLAKPTPKAETTTKQPA 300
Db 191 VETKETTTTINKQTSSTGKEKTSIAKETQSIKTSIAKOLAPTSKVLAKPTPKAETTTKQPA 250
QY 301 LTTPEKPTPTPKPEASTTKEPTPTTIKSAPTTPKPEAPPTTKSAPTTPKPEAPTTTKE 360
Db 251 LTTPEKPTPTPKPEASTTKEPTPTTIKSAPTTPKPEAPPTTKSAPTTPKPEAPTTTKE 310
QY 361 PAPTTPKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 420
Db 311 PAPTTPKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 370
QY 421 PKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 480
Db 371 PKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 430
QY 481 PTTTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 540
Db 431 PTTTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 490
QY 541 PAPTTPKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 600
Db 491 PAPTTPKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 550
QY 601 TPESLAPTTPEEPTPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEE 660
Db 551 TPESLAPTTPEEPTPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEE 610
QY 661 TAPTTPKGTATTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 720
Db 611 TAPTTPKGTATTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 670
QY 721 EPAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 780
Db 671 EPAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 730
QY 781 KETAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 840
Db 731 KETAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 790
QY 841 PTPKALENSKPEGVPTTKTAAKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 900
Db 791 PTPKALENSKPEGVPTTKTAAKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 850
QY 901 TTEKTTESKITATTTQVTSSTTQDTPPKIITLKTTLAPKVTTTKTITTEIMNKPEE 960
Db 851 TTEKTTESKITATTTQVTSSTTQDTPPKIITLKTTLAPKVTTTKTITTEIMNKPEE 910
QY 961 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMTMPRVKPKTTPPRKMTSTMPBLN 1020
Db 911 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMTMPRVKPKTTPPRKMTSTMPBLN 970
QY 1021 PTSRIAETAMLQTTTTPRNOTPNSKLVEVNPKESEDAGGEGTTPHMLLRPHVFMPEVTPDMD 1080
Db 971 PTSRIAETAMLQTTTTPRNOTPNSKLVEVNPKESEDAGGEGTTPHMLLRPHVFMPEVTPDMD 1030
QY 1081 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1140
Db 1031 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1090
QY 1141 RITEVWGIPTSIDTTFTRCNCEGKTFPPKDSQYWRFTNDIKDAGYFKPIFKGFGGLTQOI 1200
Db 1091 RITEVWGIPTSIDTTFTRCNCEGKTFPPKDSQYWRFTNDIKDAGYFKPIFKGFGGLTQOI 1150
QY 1201 VAALSTAKYKNWPESVYFKRGSGTQQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1260
Db 1151 VAALSTAKYKNWPESVYFKRGSGTQQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1210
QY 1261 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEKVSILWRGLPNVWVTSAISLENIRKPD 1320

Db 1211 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEKVSILWRGLPNVWVTSAISLENIRKPD 1270
QY 1321 GYDYIAFASKQYXNIDVPSRTARAITTRSGQTLSKQWYNCP 1361
Db 1271 GYDYIAFASKQYXNIDVPSRTARAITTRSGQTLSKQWYNCP 1311
RESULT 9
ADK67911
ID ADK67911 standard; protein; 1270 AA.
XX AC ADK67911;
XX DT 06-MAY-2004 (first entry)
XX Human extracellular messenger (EXMES) polypeptide.
DE Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
XX antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
KW cytostatic; gene therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..29
FT /label= Signal peptide
FT /note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29
XX according to identification method"
PN WO2004013292-A2.
XX 12-FEB-2004.
XX 30-JUL-2003; 2003WO-US024084.
XX 02-AUG-2002; 2002US-0400810P.
PR 19-SEP-2002; 2002US-0412197P.
PR 04-OCT-2002; 2002US-0416004P.
PR 08-NOV-2002; 2002US-0424862P.
XX (INCY-) INCYTE CORP.
XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;
PI Richardson TW, Emerling EW, Lindquist EA, Chawia NK, Ramkumar J;
PI Lee SY;
XX WPI; 2004-1571116/15.
DR N-PSDB; ADK67916.
XX New extracellular messengers and nucleic acids, useful for diagnosing,
PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
PT autoimmune thyroiditis.
XX Claim 59; SEQ ID NO 4; 165pp; English.
XX The present sequence is that of novel human extracellular messenger
CC (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology
CC to human megakaryocyte stimulating factor. The invention provides EXMES
CC polynucleotides and polypeptides, as well as expression vectors, host
CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
CC treating or preventing disorders associated with aberrant expression of
CC EXMES, especially autoimmune and inflammatory disorders, cell
CC proliferative disorders and endocrine disorders, e.g. adult respiratory
CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
CC parasitic, protozoal or helminthic infections, cancers, autoimmune
CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
CC Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery
CC process, including determining of efficacy, dosage, toxicity and
CC pharmacology, and for investigating the pathogenesis of diseases and
CC medical conditions.
XX
SQ Sequence 1270 AA;

Query Match		92.6%; Score 6778.9; DB 8; Length 1270;
Best Local Similarity		93.2%; Pred. No. 3.9e-163;
Matches 1269; Conservative		0; Mismatches 1; Indels 91; Gaps 2;
QY	1	MAWKLPYLLLLSVFVQVSSQDLSSCAGRGEGYSDATCNDYNCQHMECCPDF 60
DB	1	MAWKLPYLLLLSVFVQVSSQ----- 25
QY	61	KRVCTAELSCGRCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB	26	-----ELSCGRCFESFERGECDDAQCKYDKCCPDYSEFCA----- 64
QY	121	PPSGASQTIKSTTKRSPKPNKKTKKVI EEEIETEVKDNKKNRKTKKPTPKPPVDEAG 180
DB	65	-----EVDNKKNRKTKKPTPKPPVDEAG 89
QY	181	SGLDNGFKVTPDTSTQHNKVTSPKITTAKP INPRPSLPNSDTSKETS LTVNKETT 240
DB	90	SGLDNGDFKVTPTDTSTQHNKVTSPKITTAKP INPRPSLPNSDTSKETS LTVNKETT 149
QY	241	VEYKETTITNKQSTDGKEKTS AKETOS I EKTSAKOLAPTSKVLAKPTPKAETTTKGA 300
DB	150	VEYKETTITNKQSTDGKEKTS AKETOS I EKTSAKOLAPTSKVLAKPTPKAETTTKGA 209
QY	301	LTTPKEPTTTPKEPASTTPKEPTPTTIKSAPTTPKPEPAPTTTKSAPTTPKEPAPTTTKE 360
DB	210	LTTPKEPTTTPKEPASTTPKEPTPTTIKSAPTTPKPEPAPTTTKSAPTTPKEPAPTTTKE 269
QY	361	PAPTTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTTPKPEPAPTTTPKEPTPTT 420
DB	270	PAPTTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTTPKPEPAPTTTPKEPTPTT 329
QY	421	PKPEPAPTTTPKEPAPTTAPKPAPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPA 480
DB	330	PKPEPAPTTTPKEPAPTTAPKPAPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPA 389
QY	481	PTTTKSAPTTTKEPAPTTTKSAPTTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKE 540
DB	390	PTTTKSAPTTTKEPAPTTTKSAPTTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKE 449
QY	541	PAPTTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKDAPT 600
DB	450	PAPTTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKPAPT 509
QY	601	TPEELAPTTPEEPPTTPPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 660
DB	510	TPEELAPTTPEEPPTTPPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 569
QY	661	TAPTPKGATPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKPAPTTPKGTAPTPPK 720
DB	570	TAPTPKGATPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKPAPTTPKGTAPTPPK 629
QY	721	EPAPTTTPKEPAPTTPKGTAPTTLUKEPAPTTTPKPAKELAPTTTKEPTSTTSKPAPTTP 780
DB	630	EPAPTTTPKEPAPTTPKGTAPTTLUKEPAPTTTPKPAKELAPTTTKEPTSTTSKPAPTTP 689
QY	781	KETAPTTTPKEPAPTTTPKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840
DB	690	KETAPTTTPKEPAPTTTPKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 749
QY	841	PTPKALENSKEPGVPTTKTAAKPEMTTAKDKTTERDLRTTPEPTTAAPKMTKETAT 900
DB	750	PTPKALENSKEPGVPTTKTAAKPEMTTAKDKTTERDLRTTPEPTTAAPKMTKETAT 809
QY	901	TTBKTTESKITATTTQVTSSTTQDTPPKITLTKTTTLAPKVTITTKKTIITTEIMNKPEE 960

DB	810	TTEKTTESKITATTTQVTSSTTQDTPPKITLTKTTTLAPKVTITTKKTIITTEIMNKPEE 869
QY	961	TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMRVRKPKTTPTPRKMSTWPELN 1020
DB	870	TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMRVRKPKTTPTPRKMSTWPELN 929
QY	1021	PTSAIAEAMLOTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPMD 1080
DB	930	PTSAIAEAMLOTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPMD 989
QY	1081	YLPVPNGOIIINPMLSDETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPESPSPAR 1140
DB	990	YLPVPNGOIIINPMLSDETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPESPSPAR 1049
QY	1141	RITEVWGPSPIDTVTRCNCEGKTFPFDKSYWRFTNDIKDAGYPKPIFKGFGGLTQOI 1200
DB	1050	RITEVWGPSPIDTVTRCNCEGKTFPFDKSYWRFTNDIKDAGYPKPIFKGFGGLTQOI 1109
QY	1201	VAALSTAKYKNWPESVYFFKRGSSIQQYIYKQEPVQKCPGRRPALNIPYVYGETTQVRRR 1260
DB	1110	VAALSTAKYKNWPESVYFFKRGSSIQQYIYKQEPVQKCPGRRPALNIPYVYGETTQVRRR 1169
QY	1261	PERAIGSQTHIRIOXSPARLAYQDKGVHLNPKVSVILWRGLNVTSAISLNPRIKPD 1320
DB	1170	PERAIGSQTHIRIOXSPARLAYQDKGVHLNPKVSVILWRGLNVTSAISLNPRIKPD 1229
QY	1321	GYDYAFSKOQYINIDVPSRTARATITRSQGLTSKVMYNCP 1361
DB	1230	GYDYAFSKOQYINIDVPSRTARATITRSQGLTSKVMYNCP 1270
RESULT 10		
AAM24322		
ID	AAM24322	standard; protein; 1299 AA.
XX	AC	AAM24322;
XX	AC	
DT	12-OCT-2001	(first entry)
XX	DE	Human EST encoded protein SEQ ID NO: 1847.
XX	DE	
XX	KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	KW	diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
XX	XX	gene therapy; nutrition.
OS	OS	Homo sapiens.
XX	XX	
PN	PN	WO200154477-A2.
XX	XX	
PD	02-AUG-2001.	
XX	XX	25-JAN-2001; 2001WO-US002687.
XX	XX	
PR	25-JAN-2000; 2000US-00491404.	
PR	17-JUL-2000; 2000US-00617746.	
PR	03-AUG-2000; 2000US-00631451.	
PR	15-SEP-2000; 2000US-00663870.	
XX	XX	(HYSE-) HYSSEQ INC.
PA	PA	
XX	XX	
PI	PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	XX	WPI; 2001-476164/51.
DR	DR	N-PSDB; AAH98981.
XX	XX	
PT	PT	Isolated polypeptide for treatment of diseases, diagnostics, raising
PT	PT	antibodies and research use.
XX	XX	
PS	PS	Claim 20; Page 1198-1201; 1275pp; English.
XX	XX	
CC	CC	The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention
XX
SQ

Query Match 91.9%; Score 6732.7; DB 4; Length 1299;
Best Local Similarity 96.7%; Pred. No. 5.9e-162;
Matches 1252; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Qy 61 KRVTABLSCKGRCFESFERGECDCDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Db 61 KRVTABLSCKGRCFESFERGECDCDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Qy 121 PPSGASQTIKSTTKRSPPNNKTKKVIIESEIITE----- 156
Db 121 PPSGASQTIKSTTKRSPPNNKTKKVIIESEIITEHSVSENQESSSSSSSSSSSTI 180
Qy 157 -----VKDNKNRTKKKPTPKPPVVDVAGSLONGDFKVTTPDTST 197
Db 181 KIKSSKNSAANRELQKLKVDNKNRTKKKPTPKPPVVDVAGSLONGDFKVTTPDTST 240
Qy 198 TOHNKVSPTKITAKPNRPSPNPSDTSKETSIVNKETTVETKTTTTNKQTSIDG 257
Db 241 TOHNKVSPTKITAKPNRPSPNPSDTSKETSIVNKETTVETKTTTTNKQTSIDG 300
Qy 258 KEKTSASQTSIETSAKDLAPTSKVLAKPTKAEITTKGPALTTKPEPTTPPKSPAS 317
Db 301 KEKTSASQTSIETSAKDLAPTSKVLAKPTKAEITTKGPALTTKPEPTTPPKSPAS 360
Qy 318 TTPKEPTPTTKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 377
Db 361 TTPKEPTPTTKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 420
Qy 378 APPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 437
Db 421 APPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 480
Qy 438 EPAPTAPKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 497
Db 481 EPAPTAPKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 540
Qy 498 TTKGAPTTKPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 557
Db 541 TTKGAPTTKPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 600
Qy 558 APPTAPKEPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 617
Db 601 APPTAPKEPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 660
Qy 618 PEEPAPTTKKAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 677
Db 661 PEEPAPTTKKAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 720
Qy 678 APPTPKKAPKELAPTTKPEPTSTSDKAPTTKGTAPTTKPEAPTTKPEAPTTKPK 737
Db 721 APPTPKKAPKELAPTTKPEPTSTSDKAPTTKGTAPTTKPEAPTTKPEAPTTKPK 780
Qy 738 TAPPTLKPEAPTTKPKAPKELAPTTKGTSTSDKAPTTKPEAPTTKPEAPTTKPK 797
Db 781 TAPPTLKPEAPTTKPKAPKELAPTTKGTSTSDKAPTTKPEAPTTKPEAPTTKPK 840
Qy 798 KPAPTTPTPTPTTSEVSPPTTKGPTTHKSPDSTBELSAEPTPKALENSPKFPGVPT 857
Db 841 KPAPTTPTPTPTTSEVSPPTTKGPTTHKSPDSTBELSAEPTPKALENSPKFPGVPT 900

Qy 858 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAPKMTKETATTTESKITATTQV 917
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAPKMTKETATTTESKITATTQV 960
Qy 918 TSITTTQDTTFFPKITTLTKTTTLAPKVTTTXXKTTTTEIMNKPEETAKPKDRATNSKATTPK 977
Db 961 TSITTTQDTTFFPKITTLTKTTTLAPKVTTTXXKTTTTEIMNKPEETAKPKDRATNSKATTPK 1020
Qy 978 PQKETKAPKPTSTKPKTMERVRKPKTTPPRKMTSTMPBELNPTSRISAEAMLQTTTRPN 1037
Db 1021 PQKETKAPKPTSTKPKTMERVRKPKTTPPRKMTSTMPBELNPTSRISAEAMLQTTTRPN 1080
Qy 1038 QTPNSKLIVNPKSEDAGAEGETPHMLRPHVFWPEVTPDMDVLPVNPNOGIIINPMLS 1097
Db 1081 QTPNSKLIVNPKSEDAGAEGETPHMLRPHVFWPEVTPDMDVLPVNPNOGIIINPMLS 1140
Qy 1098 DETNINCKGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSIDTVFT 1157
Db 1141 DETNINCKGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSIDTVFT 1200
Qy 1158 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKWPESVY 1217
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKWPESVY 1260
Qy 1218 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGE 1252
Db 1261 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGE 1295

RESULT 11
AAB29778 standard; protein; 902 AA.
AC AAB29778;
DT 28-FEB-2001 (first entry)
DE Human MSF-derived tribonectin.
KW Human tribonectin; MSF; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
OS Homo sapiens.
PN WO200064930-A2.
PD 02-NOV-2000.
PF 24-APR-2000; 2000WO-US010953.
PR 23-APR-1999; 99US-00298970.
PA (RHOD-) RHODE ISLAND HOSPITAL LIPESPAN PARTNER.
PI Jay GD;
DR WPI; 2001-024673/03.
PT Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
PS Disclosure; Fig 1; 47pp; English.
CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonectin; a biocompatible composition comprising a human

QY 554 TKKPAATAPKEPAPTTPEKAPTTTPKKTTPTTPEKLAAPTTPEKAPTTPEELAPTTPEEP 613
 XX |||||
 Db 1 TKKPAATAPKEPAPTTPEKAPTTTPKKTTPTTPEKLAAPTTPEKAPTTPEELAPTTPEEP 60
 QY 614 TPTTPEEPAPTTPKAAANPTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAATTPKGTAPTT 673
 XX |||||
 Db 61 TPTTPEEPAPTTPKAAANPTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAATTPKGTAPTT 120
 QY 674 LKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTT 733
 XX |||||
 Db 121 LKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTT 180
 QY 734 TPKGTAATTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAATTPKEPAP 793
 XX |||||
 Db 181 TPKGTAATTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAATTPKEPAP 240
 QY 794 TTPKKAPAPTTPEPPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKEP 853
 XX |||||
 Db 241 TTPKKAPAPTTPEPPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKEP 300
 QY 854 GVTPTTKPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTTESKITAT 913
 XX |||||
 Db 301 GVTPTTKPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTTESKITAT 360
 QY 914 TTQVSTTTQDTPPEKITTLLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKA 973
 XX |||||
 Db 361 TTQVSTTTQDTPPEKITTLLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKA 420
 QY 974 TTPKPKPTKAPKKPTSTKKPKTMRVRKPKTTTPPKMTSTMPELNPTSRIAEAMLQTT 1033
 XX |||||
 Db 421 TTPKPKPTKAPKKPTSTKKPKTMRVRKPKTTTPPKMTSTMPELNPTSRIAEAMLQTT 480
 QY 1034 TRNQTPNSKLVENPKSEBAGGEGTPEHMLLRPHVFMPEVTPDMDYLRVFNQGIIN 1093
 XX |||||
 Db 481 TRNQTPNSKLVENPKSEBAGGEGTPEHMLLRPHVFMPEVTPDMDYLRVFNQGIIN 540
 QY 1094 PMLSDETNICN 1104
 XX |||||
 Db 541 PMLSDETNICN 551

RESULT 13

ABU53252

ID ABU53252 standard; protein; 546 AA.

XX

AC ABU53252;

XX

DT 14-APR-2003 (first entry)

XX

DE Human testes-derived DKFZp385_4019 homologue #1.

XX

KW Human; gene therapy; vaccine; disease treatment; detection.

XX

OS Homo sapiens.

XX

PN W0200112659-A2.

XX

FD 22-FEB-2001.

XX

XX 18-AUG-2000; 2000WO-IB001496.

XX

XX 18-AUG-1999; 99US-0149499P.

PR

XX 28-SEP-1999; 99US-0156503P.

XX

FA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX

PI Wiemann S;

XX

XX WPI; 2001-327840/34.

XX

PT Nucleic acids having the sequences of clones isolated from libraries of

XX

PT different human tissues, useful in recombinant DNA methodologies.

XX

PS Example III; Page 892; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention

XX Sequence 546 AA;

Query Match 39.9%; Score 2920; DB 4; Length 546;
 Best Local Similarity 100.0%; Pred. No. 4.8e-66;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 KKPAATTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTTSAPTITTKAPTTTKSAPT 504
 XX |||||
 Db 1 KKPAATTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTTSAPTITTKAPTTTKSAPT 60
 QY 505 TPKEPSPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKKAPAPAPKE 564
 XX |||||
 Db 61 TPKEPSPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKKAPAPAPKE 120
 QY 565 PAPTTPKETAATTPKKLTTPTPKLAATTPPEKAPAPTTPEELAPTTPEEPPTTPEEPAPT 624
 XX |||||
 Db 121 PAPTTPKETAATTPKKLTTPTPKLAATTPPEKAPAPTTPEELAPTTPEEPPTTPEEPAPT 180
 QY 625 TPXAAADNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTILKEPAPTTPKK 684
 XX |||||
 Db 181 TPXAAADNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTILKEPAPTTPKK 240
 QY 685 PAPKELAPTTTKEPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLK 744
 XX |||||
 Db 241 PAPKELAPTTTKEPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLK 300
 QY 745 EPAPTTTPKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAPTTTPKEPAPTTTPKAPAPTTTP 804
 XX |||||
 Db 301 EPAPTTTPKAPKELAPTTTKEPTSTSDKAPAPTTTPKETAPTTTPKEPAPTTTPKAPAPTTTP 360
 QY 805 ETPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKEPGVPTTKTAAAT 864
 XX |||||
 Db 361 ETPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKEPGVPTTKTAAAT 420
 QY 865 KPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTTESKITATTTQVSTTTQD 924
 XX |||||
 Db 421 KPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTTESKITATTTQVSTTTQD 480
 QY 925 TTPFKITTLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 984
 XX |||||
 Db 481 TTPFKITTLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540
 QY 985 PKKPTS 990
 XX |||||
 Db 541 PKKPTS 546

RESULT 14

AAO18834

ID AAO18834 standard; protein; 538 AA.

XX

AC AAO18834;

XX

XX 29-OCT-2002 (first entry)

XX

3' cartilage superficial zone protein coding sequence encoded protein.

XX

SZP; superficial zone protein; cartilage; lubrication; human;

XX

degenerative joint condition; arthritis; osteoporosis; trauma; CACP;

XX

chondroitin sulphate substitution consensus; antiarthritic;

KW antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;
 KW immunosuppressive.
 XX
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Misc-difference 114 /note= "encoded by ACTACT"
 FT
 FT
 FT
 XX WO200262847-A2.
 XX
 XX 15-AUG-2002.
 PD
 PD 31-DEC-2001; 2001WO-US050379.
 PF
 PF 29-DEC-2000; 2000US-0258920P.
 PR
 XX (GLAXO) GLAXO GROUP LTD.
 PA (RUSH-) RUSH PRESBYTERIAN ST LUKE MEDICAL CENT.
 XX
 XX Hutchins JT, Kuettnner KE, Schmid TM, Schumacher BL, Su J;
 PI Dixon EP;
 XX
 XX WPI; 2002-636585/68.
 DR N-PSDB; AAL49079.
 DR
 XX New purified superficial zone protein (SZP) polypeptides, useful for
 XX treating degenerative joint conditions, e.g. osteoarthritis, rheumatoid
 PT arthritis, gout, spondylarthritis, synovitis, tendonitis, lupus, or
 PT osteoporosis.
 PT
 XX
 XX Claim 59; Page 86-87; 89pp; English.
 PS
 XX
 XX The present invention provides the protein and coding sequences of human
 CC superficial zone protein (SZP). The protein is involved in the
 CC lubrication of joints, and the sequences can be used in the treatment of
 CC degenerative joint conditions or to delay symptoms of a degenerative
 CC joint condition, e.g. osteoarthritis, rheumatoid arthritis, gout,
 CC psoriatic arthritis, reactive arthritis, viral or post viral arthritis,
 CC spondylarthritis, juvenile arthritis, synovitis, tendonitis, systemic
 CC lupus erythematosus, CACP, osteoporosis or trauma. The present sequence
 CC is the protein encoded by the human 3' cartilage SZP cDNA
 XX
 XX Sequence 538 AA;
 PS
 PS Query Match 38.9%; Score 2850.9; DB 5; Length 538;
 PS Best Local Similarity 99.8%; Pred. No. 2.6e-64;
 PS Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 823 PTTIHKSPDESTPELSAEPFKALENSPKPEPGVPTTKPAATKPEMTTAKDKTTERDLR 882
 Db 1 PTTIHKSPDESTPELSAEPFKALENSPKPEPGVPTTKPAATKPEMTTAKDKTTERDLR 60
 QY 883 TPTETTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLKTTLAPKV 942
 Db 61 TPTETTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLKTTLAPKV 119
 QY 943 TTTTKKITTITEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTKKPKTMRVRK 1002
 Db 120 TTTTKKITTITEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTKKPKTMRVRK 179
 QY 1003 PKTTPPRKMTSTWPELNPSTRIAEAMLOTTTPRNPQNSKLVENPKPSDAGAGETP 1062
 Db 180 PKTTPPRKMTSTWPELNPSTRIAEAMLOTTTPRNPQNSKLVENPKPSDAGAGETP 239
 QY 1063 HMLLRPHVFPEVTPDMDYLPVRPNQGIINPMLSDETNICNGKVPDGLTTLRNGTLVAF 1122
 Db 240 HMLLRPHVFPEVTPDMDYLPVRPNQGIINPMLSDETNICNGKVPDGLTTLRNGTLVAF 299
 QY 1123 RGHYFWMLSFSPSPARRITEVWGPISPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKD 1182
 Db 300 RGHYFWMLSFSPSPARRITEVWGPISPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKD 359

QY 1183 AGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVYFFKRGSIQQYIYKQEPVQKCPGRR 1242
 Db |||||
 QY 360 AGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVYFFKRGSIQQYIYKQEPVQKCPGRR 419
 Db |||||
 QY 1243 PALNYPVYGEMTQVRRRRFERRAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRG 1302
 Db |||||
 QY 420 PALNYPVYGEMTQVRRRRFERRAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRG 479
 Db |||||
 QY 1303 LPNVVTSALSLPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTRSGQTLSKWVNCP 1361
 Db |||||
 QY 480 LPNVVTSALSLPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTRSGQTLSKWVNCP 538
 Db |||||
 RESULT 15
 ABUS3254
 ID ABUS3254 standard; protein; 513 AA.
 XX
 XX AC ABUS3254;
 XX
 XX 14-APR-2003 (first entry)
 XX
 XX Human testes-derived DKFZphtes3_4019 homologue #3.
 DE
 XX Human; gene therapy; vaccine; disease treatment; detection.
 KW
 KW Homo sapiens.
 OS
 XX WO200112659-A2.
 PN
 XX 22-FEB-2001.
 PD
 XX
 XX 18-AUG-2000; 2000WO-IB001496.
 PF
 XX
 XX 18-AUG-1999; 99US-0149499P.
 PR
 XX 28-SEP-1999; 99US-0156503P.
 PR
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 PA
 XX Wiemann S;
 PI
 XX WPI; 2001-327840/34.
 DR
 XX Nucleic acids having the sequences of clones isolated from libraries of
 PT different human tissues, useful in recombinant DNA methodologies.
 PT
 XX Example III; Page 893; 1095pp; English.
 PS
 XX This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX
 XX Sequence 513 AA;
 PS
 PS Query Match 37.6%; Score 2757; DB 4; Length 513;
 PS Best Local Similarity 100.0%; Pred. No. 5.8e-62;
 PS Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 164 RTKKXPTPKPPVVDEAGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINPESLPP 223
 Db 1 RTKKXPTPKPPVVDEAGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINPESLPP 60
 QY 224 NSDTSKTSLSLVNKKETTVETKETTITNKQSTDKCKTTSKAKTQSIKTSKDLAPTSK 283
 Db 61 NSDTSKTSLSLVNKKETTVETKETTITNKQSTDKCKTTSKAKTQSIKTSKDLAPTSK 120
 QY 284 VLAKPTPKAETTTKGPALTPPKPTPTPKPEASTTPKPTTTIKSAPTTPKEPAPTTT 343

```
Db      121  VLAKTPKAEITTYKGPALTTPKEPTTTPKEPASTTPKBPPTTIKSAPTTPKEPAPTTT 180
QY      344  KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKPDAPT 403
Db      181  KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKPDAPT 240
QY      404  TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 463
Db      241  TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 300
QY      464  APITTKEPSPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKEPAPTTTPK 523
Db      301  APITTKEPSPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKSAPTTPKEPAPTTTPKEPAPTTTPK 360
QY      524  EPAPTTPKPDAPTTPKEPAPTTTPKEPAPTTTPKPDAPTTPKPDAPTTPKPDAPTTPKPDAPT 583
Db      361  EPAPTTPKPDAPTTPKEPAPTTTPKEPAPTTTPKPDAPTTPKPDAPTTPKPDAPTTPKPDAPT 420
QY      584  TTPEKLAAPTTPKEPAPTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTP 643
Db      421  TTPEKLAAPTTPKEPAPTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTP 480
QY      644  EPAPTTPKEPAPTTPKETAPTTPKGTAAPTTLKE 676
Db      481  EPAPTTPKEPAPTTPKETAPTTPKGTAAPTTLKE 513
```

Search completed: October 13, 2004, 11:37:22
Job time : 116.728 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 23.6582 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKLPYVLLLSVFIQ.....ARAITRSQGTLKSVWYVNC 1361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES												
Result No.	Score	Query Match	Length	ID	Description							
1	7323	100.0	1361	4	US-07-757-022B-40	Sequence 40, Appl						
2	7308.7	99.8	1404	4	US-07-757-022B-2	Sequence 2, Appli						
3	7308.7	99.8	1404	4	US-07-757-022B-62	Sequence 62, Appl						
4	7308.7	99.8	1404	4	US-09-298-970A-1	Sequence 1, Appli						
5	7293.7	99.6	1404	4	US-10-164-595-78	Sequence 78, Appl						
6	7084.3	96.7	1354	4	US-07-757-022B-48	Sequence 48, Appl						
7	7058.9	96.4	1320	4	US-07-757-022B-46	Sequence 46, Appl						
8	7058.9	96.4	1320	4	US-07-757-022B-60	Sequence 60, Appl						
9	7049	96.3	1311	4	US-07-757-022B-42	Sequence 42, Appl						
10	7044.6	96.2	1363	4	US-07-757-022B-52	Sequence 52, Appl						
11	7043.9	96.2	1320	4	US-10-164-595-58	Sequence 58, Appl						
12	6836.3	93.4	1314	4	US-07-757-022B-50	Sequence 50, Appl						
13	6820.2	93.1	1313	4	US-07-757-022B-142	Sequence 142, App						
14	6784.9	92.7	1270	4	US-07-757-022B-44	Sequence 44, Appl						
15	5867.7	80.1	1140	4	US-07-757-022B-104	Sequence 104, App						
16	5485.7	74.9	1038	4	US-07-757-022B-74	Sequence 74, Appl						
17	5472	74.7	1022	4	US-07-757-022B-84	Sequence 84, Appl						
18	5379.2	73.5	1049	4	US-07-757-022B-58	Sequence 58, Appl						
19	5011	68.4	941	4	US-07-757-022B-14	Sequence 14, Appl						
20	2246.2	30.7	463	4	US-07-757-022B-54	Sequence 54, Appl						
21	1991.2	27.2	423	4	US-07-757-022B-66	Sequence 66, Appl						
22	1982.1	27.1	422	4	US-07-757-022B-68	Sequence 68, Appl						
23	1723.1	23.5	372	4	US-07-757-022B-64	Sequence 64, Appl						
24	1413.6	19.3	5179	4	US-09-538-032-1258	Sequence 1258, Ap						
25	1295.5	17.7	236	4	US-07-757-022B-70	Sequence 70, Appl						
26	1214.6	16.6	237	4	US-07-757-022B-72	Sequence 72, Appl						
27	1189.7	16.2	8991	4	US-08-714-741-32	Sequence 32, Appl						

ALIGNMENTS

RESULT 1
US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID

Sequence 116, App
Sequence 136, App
Sequence 96, Appl
Sequence 30, Appl
Sequence 32, Appl
Sequence 94, Appl
Sequence 132, App
Sequence 92, Appl
Sequence 88, Appl
Sequence 1, Appli
Sequence 106, App
Sequence 118, App
Sequence 102, App
Sequence 114, App
Sequence 2, Appli
Sequence 98, Appl
Sequence 22, Appl

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
US-07-757-022B-40

Query Match      100.0%; Score 7323; DB 4; Length 1361;
Best Local Similarity 100.0%; Pred. No. 2e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNHMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNHMECCPDF 60

Qy 61 KRVTAEISCKGRCFESFERGECDDAQCKKYDKCCPDYFSCAEVHNPTSPSSKKAP 120
Db 61 KRVTAEISCKGRCFESFERGECDDAQCKKYDKCCPDYFSCAEVHNPTSPSSKKAP 120

Qy 121 PPGASQTIKSTTKRSKPKPNKKTKVIESEETEVKDNKNKTKKKPTPKPPVDEAG 180
Db 121 PPGASQTIKSTTKRSKPKPNKKTKVIESEETEVKDNKNKTKKKPTPKPPVDEAG 180

Qy 181 SGLDNGDFKVTTPDTSTTOHNKVSPTKITTAKDINRPRLPPNSDTSKETSIVNKETT 240
Db 181 SGLDNGDFKVTTPDTSTTOHNKVSPTKITTAKDINRPRLPPNSDTSKETSIVNKETT 240

Qy 241 VETKETTITNKQTSDDGKEKTTSAKETQSTKTSKVLAKPTPKAETTTKGPA 300
Db 241 VETKETTITNKQTSDDGKEKTTSAKETQSTKTSKVLAKPTPKAETTTKGPA 300

Qy 301 LTTPKPTPTTKEPASTTKEPTPTTKSAPTTKPEPAPTTTKSAPTTKPEAPTTTKE 360
Db 301 LTTPKPTPTTKEPASTTKEPTPTTKSAPTTKPEPAPTTTKSAPTTKPEAPTTTKE 360

Qy 361 PAPTTPKEPAPTTTKBAPTTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTKEPTPT 420
Db 361 PAPTTPKEPAPTTTKBAPTTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTKEPTPT 420

Qy 421 PKBAPTTKEPAPTTKPEAPTAPEKAPTTKPEPAPTTKPEPAPTTTKESPTTKPEA 480
Db 421 PKBAPTTKEPAPTTKPEAPTAPEKAPTTKPEPAPTTKPEPAPTTTKESPTTKPEA 480

Qy 481 PTTTKSAPTTTKEPAPTTTKSAPTTKPEPSTTTKPEPAPTTKPEPAPTTPKKAPTTPK 540
Db 481 PTTTKSAPTTTKEPAPTTTKSAPTTKPEPSTTTKPEPAPTTKPEPAPTTPKKAPTTPK 540

Qy 541 PAPTTPKEPAPTTTKBAPTAPEKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKEPAPT 600
Db 541 PAPTTPKEPAPTTTKBAPTAPEKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKEPAPT 600

Qy 601 TPEELAPTTPEPTPTTKEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKE 660
Db 601 TPEELAPTTPEPTPTTKEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKE 660

Qy 661 TAPTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKGTAPTPK 720
Db 661 TAPTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKGTAPTPK 720

Qy 721 EPAPTTPKEPAPTTKGTAPTTLKEPAPTTKPEPAPTTKPELAPTTTKGPTSTSDKAPTP 780
Db 721 EPAPTTPKEPAPTTKGTAPTTLKEPAPTTKPEPAPTTKPELAPTTTKGPTSTSDKAPTP 780

Qy 781 KETAPTTPKBAPTTPKKAPPTTPEPTPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840
Db 781 KETAPTTPKBAPTTPKKAPPTTPEPTPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840

Qy 841 PTPKALENSPKBPGVPTTKTBAATKPEMTTAKKTTTERDLRTPPTTAAAPKMTKETAT 900
Db 841 PTPKALENSPKBPGVPTTKTBAATKPEMTTAKKTTTERDLRTPPTTAAAPKMTKETAT 900

Qy 901 TTEKTTESKITAATTTQVTSITTTQDTPPKITITLKTITLAPKVTTTKITITTEIMNKP 960
Db 901 TTEKTTESKITAATTTQVTSITTTQDTPPKITITLKTITLAPKVTTTKITITTEIMNKP 960

Qy 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTPRKWTSTMPBLN 1020
Db 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTPRKWTSTMPBLN 1020

RESULT 2
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
```


TELFAAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-2

Query Match 99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLILLLSVFVIQVSSODLSSCAGRGEGYSDRATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLILLLSVFVIQVSSODLSSCAGRGEGYSDRATCNCYNCOHYMECCPDF 60

QY 61 KRVCCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYEFCAEVAHNPTSPSSKKAP 120
Db 61 KRVCCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYEFCAEVAHNPTSPSSKKAP 120

QY 121 PPSCASOTIKSTTKRSPKPNKKTKKVIIEEITE----- 156
Db 121 PPSCASOTIKSTTKRSPKPNKKTKKVIIEEITE----- 156

QY 157 -----VKDNKNRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 180
Db 157 -----VKDNKNRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 180

QY 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 240
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTKPPVVDGAGSLDNGDFKVTTPDTST 240

QY 198 TOHNKYSTPKITAKPINSRPSLPNSDTSKETSLSLVNKEITVETKETTITNNKQTSIDG 257
Db 198 TOHNKYSTPKITAKPINSRPSLPNSDTSKETSLSLVNKEITVETKETTITNNKQTSIDG 257

QY 241 TOHNKYSTPKITAKPINSRPSLPNSDTSKETSLSLVNKEITVETKETTITNNKQTSIDG 300
Db 241 TOHNKYSTPKITAKPINSRPSLPNSDTSKETSLSLVNKEITVETKETTITNNKQTSIDG 300

QY 258 KEKTTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 317
Db 258 KEKTTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 317

QY 301 KEKTTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 360
Db 301 KEKTTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 360

QY 318 TTPEPTTPPKBPASPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 377
Db 318 TTPEPTTPPKBPASPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 377

QY 361 TTPEPTTPPKBPASPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420
Db 361 TTPEPTTPPKBPASPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420

QY 378 APPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 437
Db 378 APPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 437

QY 421 APPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 480
Db 421 APPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 480

QY 438 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 497
Db 438 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 497

QY 481 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540
Db 481 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540

QY 498 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 557
Db 498 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 557

QY 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
Db 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600

QY 558 APAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 617
Db 558 APAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 617

QY 601 APAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 660
Db 601 APAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 660

QY 618 PEPAPPTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 677
Db 618 PEPAPPTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 677

QY 661 PEPAPPTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720
Db 661 PEPAPPTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720

QY 678 APPTTKKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 737
Db 678 APPTTKKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 737

QY 721 APPTTKKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 780
Db 721 APPTTKKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 780

QY 738 TAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 797
Db 738 TAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 797

QY 781 TAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 840
Db 781 TAPTTKBPAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 840

QY 798 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALNSPKBPGVPT 857
Db 798 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALNSPKBPGVPT 857

QY 841 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALNSPKBPGVPT 900
Db 841 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALNSPKBPGVPT 900

RESULT 3
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-62

Query Match          99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY      1  MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
Db
QY      61  KRVTAEISCKGRCFESFERGREGCCDAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
Db
QY      121  PPGASQTIKSTTRKSPKPPNKKTKKVIIESEITE----- 156
Db
QY      121  PPGASQTIKSTTRKSPKPPNKKTKKVIIESEITEHSVSENESSSSSSSSSSSTI 180
QY      157  -----VNDKNKRTKKPTPKPVVDVDSGLDNGDFKVTTPDTST 197
Db
QY      181  KIKSSKNSAANRELQKLLKVDKNKKNRTKKPTPKPVVDVDSGLDNGDFKVTTPDTST 240
QY      198  TQHNKVSTSPKITAKPINRPSLPNDSKTSKSLTVNKEITVETKETTITNKQSTDG 257
Db
QY      241  TQHNKVSTSPKITAKPINRPSLPNDSKTSKSLTVNKEITVETKETTITNKQSTDG 300
QY      258  KEKTSKAKTOSIEKTSKADLAPTSKVLAKEPTPKAETTKGPAITTPKEPTPTPKEPAS 317
Db
QY      301  KEKTSKAKTOSIEKTSKADLAPTSKVLAKEPTPKAETTKGPAITTPKEPTPTPKEPAS 360
QY      318  TTPKEPTPTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 377
Db
QY      361  TTPKEPTPTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY      378  APTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 437
Db
QY      421  APTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 480
QY      438  EPAPTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 497
Db
QY      481  EPAPTAPKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540
QY      498  TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 557
Db
QY      541  TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600
QY      558  APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 617
Db
QY      601  APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 660
QY      618  PEEPAPTTTPKAAAPNTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 677
Db
QY      661  PEEPAPTTTPKAAAPNTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720
QY      678  APTTPKAPKAPKELAPTTTKEPTSTSDKAPTTTPKGATPTTKEPAPTTTKEPAPTTTKEP 737
Db
QY      721  APTTPKAPKAPKELAPTTTKEPTSTSDKAPTTTPKGATPTTKEPAPTTTKEPAPTTTKEP 780
```

```
QY      738  TAPTTLKEPAPTTTPKKAPKELAPTTTGGPTSTTSKDPAPTTPKETAPTTTPKEPAPTTPK 797
Db
QY      781  TAPTTLKEPAPTTTPKKAPKELAPTTTGGPTSTTSKDPAPTTPKETAPTTTPKEPAPTTPK 840
Db
QY      798  KPAPTTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAETTPKALENSPKPEGVPT 857
Db
QY      841  KPAPTTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAETTPKALENSPKPEGVPT 900
QY      858  TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKPMKTKETATTTTETKTESKITATTQV 917
Db
QY      901  TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKPMKTKETATTTTETKTESKITATTQV 960
QY      918  TSTTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 977
Db
QY      961  TSTTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY      978  PQKPTKAPKPTSTKPKTTPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1037
Db
QY      1021  PQKPTKAPKPTSTKPKTTPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080
QY      1038  QTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1097
Db
QY      1081  QTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
QY      1098  DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1157
Db
QY      1141  DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1200
QY      1158  RCNCEGHTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPSVY 1217
Db
QY      1201  RCNCEGHTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPSVY 1260
QY      1218  FFKEGGSIQYIYKQEPVQKCPGRPALNVYVYEMTQVRRRRRERAIQSOTHTIRIQY 1277
Db
QY      1261  FFKEGGSIQYIYKQEPVQKCPGRPALNVYVYEMTQVRRRRRERAIQSOTHTIRIQY 1320
QY      1278  SPARLAYQDKGVLHNEVKVSIWRLGNVVTSAISLNPKNKPDGDYDYAFSKDOYVNDV 1337
Db
QY      1321  SPARLAYQDKGVLHNEVKVSIWRLGNVVTSAISLNPKNKPDGDYDYAFSKDOYVNDV 1380
QY      1338  PSRTARAITTRSGQTLSKVMYNCP 1361
Db
QY      1381  PSRTARAITTRSGQTLSKVMYNCP 1404
```

```
RESULT 4
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1
```

```
Query Match          99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY      1  MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
Db
QY      1  MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
QY      61  KRVTAEISCKGRCFESFERGREGCCDAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
```

Db	61	KRVCTAELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNTPSPSSKKAP	120
Qy	121	PPSGASQTIKSTTKRSPKPNKKTKKVI ESEITE-----	156
Db	121	PPSGASQTIKSTTKRSPKPNKKTKKVI ESEITE-----	180
Qy	157	-----VKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDST	197
Db	181	KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDST	240
Qy	198	TOHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQTSIDG	257
Db	241	TOHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQTSIDG	300
Qy	258	KEKTTSAKETQSIEKTSKADLAPTSKVLAETPKAETTTKGPALTTTPKEPTTTPKEPAS	317
Db	301	KEKTTSAKETQSIEKTSKADLAPTSKVLAETPKAETTTKGPALTTTPKEPTTTPKEPAS	360
Qy	318	TTTKEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTKEP	377
Db	361	TTTKEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTKEP	420
Qy	378	APTTTKSAPTTTPKEPATTTPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTPK	437
Db	421	APTTTKSAPTTTPKEPATTTPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTPK	480
Qy	438	EPAPTAPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT	497
Db	481	EPAPTAPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT	540
Qy	498	TTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKP	557
Db	541	TTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKP	600
Qy	558	APTAPKEPATTTPKETAPTTPKKLTPPTPEKLAETTPKEPATTTPEELAPTTPEPTPT	617
Db	601	APTAPKEPATTTPKETAPTTPKKLTPPTPEKLAETTPKEPATTTPEELAPTTPEPTPT	660
Qy	618	PEEPAPTTPKAAANPTPKEPATTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT	677
Db	661	PEEPAPTTPKAAANPTPKEPATTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT	720
Qy	678	APTTPKKPAKELAPTTTKGPTSTTSKPAETTPKGTAPTTTPKEPATTTTPKEPATTPKG	737
Db	721	APTTPKKPAKELAPTTTKGPTSTTSKPAETTPKGTAPTTTPKEPATTTTPKEPATTPKG	780
Qy	738	TAPTTLKEPATTPKKPAKELAPTTTKGPTSTTSKPAETTPKGTAPTTTPKEPATTPPK	797
Db	781	TAPTTLKEPATTPKKPAKELAPTTTKGPTSTTSKPAETTPKGTAPTTTPKEPATTPPK	840
Qy	798	KPAPTTPETPPPTSEVSTPTTKETPTTIHKSDESTPELSAETTPKALENSPKBPVPT	857
Db	841	KPAPTTPETPPPTSEVSTPTTKETPTTIHKSDESTPELSAETTPKALENSPKBPVPT	900
Qy	858	TKTPAAKPEMTITAKDKTTERDLRTTPETTTAAPKMTKETATTTKTESKITATTQV	917
Db	901	TKTPAAKPEMTITAKDKTTERDLRTTPETTTAAPKMTKETATTTKTESKITATTQV	960
Qy	918	TSITTTQDTPPKKLTTLKTTLAPKVTTKKIITTEIMNKPEETPAKDRATNSKATPK	977
Db	961	TSITTTQDTPPKKLTTLKTTLAPKVTTKKIITTEIMNKPEETPAKDRATNSKATPK	1020
Qy	978	POKPTKAPKPTSTKPKKTPRVRKPKTTTPKMTSTMPELNPTTSIAEAMLTQTTTRPN	1037
Db	1021	POKPTKAPKPTSTKPKKTPRVRKPKTTTPKMTSTMPELNPTTSIAEAMLTQTTTRPN	1080
Qy	1038	QTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPDMYLPVVPNOGIIINPMLS	1097
Db	1081	QTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPDMYLPVVPNOGIIINPMLS	1140
Qy	1098	DETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGPSPIDTFTV	1157

Db	1141	DETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGPSPIDTFTV	1200
Qy	1158	RCNCEGKTFPKDSQYWRFTNDIKDAGYPKPIFKGFGGLTQGI VAALSTAKYKNWPESVY	1217
Db	1201	RCNCEGKTFPKDSQYWRFTNDIKDAGYPKPIFKGFGGLTQGI VAALSTAKYKNWPESVY	1260
Qy	1218	FFKGGGSIQOYIYKQEPVQKCPGRPALNYPVYGMQVRRRRFERAIGPSQTHIRIQY	1277
Db	1261	FFKGGGSIQOYIYKQEPVQKCPGRPALNYPVYGMQVRRRRFERAIGPSQTHIRIQY	1320
Qy	1278	SPARLAVQDKGLVHNEVKVSIILWGLPNVVTSAISLNRKPCGVYVYAPSKDOYXNIDV	1337
Db	1321	SPARLAVQDKGLVHNEVKVSIILWGLPNVVTSAISLNRKPCGVYVYAPSKDOYXNIDV	1380
Qy	1338	PSRTARAITRRSGQTLISKVWYNCP	1361
Db	1381	PSRTARAITRRSGQTLISKVWYNCP	1404
RESULT 5			
US-10-164-595-78			
; Sequence 78, Application US/10164595			
; Patent No. 6657054			
; GENERAL INFORMATION:			
; APPLICANT: Origene Technologies, Inc			
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides			
; FILE REFERENCE: IU 103 R1			
; CURRENT APPLICATION NUMBER: US/10/164,595			
; CURRENT FILING DATE: 2002-06-10			
; NUMBER OF SEQ ID NOS: 80			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 78			
; LENGTH: 1404			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-164-595-78			
Query Match 39.6%; Score 7293.7; DB 4; Length 1404;			
Best Local Similarity 96.7%; Pred. No. 1.4e-138;			
Matches 1358; Conservative 0; Mismatches 3; Indels 43; Gaps 1;			
Qy	1	MAWKTLPIYLLLLSVFVIOQVSSQDLSSCAGCGEGYSRDATCNCDCYNCOHYMECCPDF	60
Db	1	MAWKTLPIYLLLLSVFVIOQVSSQDLSSCAGCGEGYSRDATCNCDCYNCOHYMECCPDF	60
Qy	61	KRVCTAELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNTPSPSSKKAP	120
Db	61	KRVCTAELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNTPSPSSKKAP	120
Qy	121	PPSGASQTIKSTTKRSPKPNKKTKKVI ESEITE-----	156
Db	121	PPSGASQTIKSTTKRSPKPNKKTKKVI ESEITE-----	180
Qy	157	-----VKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDST	197
Db	181	KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDST	240
Qy	198	TOHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQTSIDG	257
Db	241	TOHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQTSIDG	300
Qy	258	KEKTTSAKETQSIEKTSKADLAPTSKVLAETPKAETTTKGPALTTTPKEPTTTPKEPAS	317
Db	301	KEKTTSAKETQSIEKTSKADLAPTSKVLAETPKAETTTKGPALTTTPKEPTTTPKEPAS	360
Qy	318	TTTKEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTKEP	377
Db	361	TTTKEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTKEP	420
Qy	378	APTTTKSAPTTTPKEPATTTPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTPK	437
Db	421	APTTTKSAPTTTPKEPATTTPKKPAETTPKEPATTTTPKEPATTTTPKEPATTTPK	480

Db 200 VETKEITTTNKQSTDCGKEKITSKETSQIEKTSKOLAPTSKVLAKPTPKAETTTKGPA 259
QY 301 LTTPEKPTPTPKPEASTTKPEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKPEAPTTTKE 360
Db 260 LTTPEKPTPTPKPEASTTKPEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKPEAPTTTKE 319
QY 361 PAPTTPKEPAPTITTKSAPTTTKSAPTTTKPEPAPTITTKKAPPTPKPEAPTTTKEPTPTT 420
Db 320 PAPTTPKEPAPTITTKSAPTTTKSAPTTTKPEPAPTITTKKAPPTPKPEAPTTTKEPTPTT 379
QY 421 PKEPAPTTPKEPAPTTPKEPAPTAPKAPPTTPKEPAPTTPKEPAPTITTKKPSPTTPKEPA 480
Db 380 PKEPAPTTPKEPAPTTPKEPAPTAPKAPPTTPKEPAPTTPKEPAPTITTKKPSPTTPKEPA 439
QY 481 PTTTKSAPTTTKPEPAPTITTKSAPTTTKPEPSPSTTTTKEPAPTTPKEPAPTTPKPKAPPTTPKE 540
Db 440 PTTTKSAPTTTKPEPAPTITTKSAPTTTKPEPSPSTTTTKEPAPTTPKEPAPTTPKPKAPPTTPKE 499
QY 541 PAPTTPKEPAPTITTKKAPTAPEPAPTTPKETAAPTTPKLTPTTPKLAAPTTPKEAPT 600
Db 500 PAPTTPKEPAPTITTKKAPTAPEPAPTTPKETAAPTTPKLTPTTPKLAAPTTPKEAPT 559
QY 601 TPEELAPTTPEEPTPTTPEEAPPTTPKAAAPNTPEPAPTTPKEPAPTTPKEPAPTTPKE 660
Db 560 TPEELAPTTPEEPTPTTPEEAPPTTPKAAAPNTPEPAPTTPKEPAPTTPKEPAPTTPKE 619
QY 661 TAPPTPKGTAPTTTKEPAPTTPKPAPELAPTTTKEPTSTTSKPAAPTTPKGTAPPTPK 720
Db 620 TAPPTPKGTAPTTTKEPAPTTPKPAPELAPTTTKEPTSTTSKPAAPTTPKGTAPPTPK 679
QY 721 EPAPTTPKEPAPTTPKGTAPTTTKEPAPTTPKPAPELAPTTTKEPTSTTSKPAAPTTPK 780
Db 680 EPAPTTPKEPAPTTPKGTAPTTTKEPAPTTPKPAPELAPTTTKEPTSTTSKPAAPTTPK 739
QY 781 KETAPTTPKEPAPTTPKPAAPTTPETPTPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840
Db 740 KETAPTTPKEPAPTTPKPAAPTTPETPTPTTSEVSTPTTKEPTTIHKSPESTPELSAE 799
QY 841 PTPKALENSKPEGVPTTKTAPAKPTKAPKPTSTKPKTTPRVRKPTTTPRXTMTSTPELN 1020
Db 800 PTPKALENSKPEGVPTTKTAPAKPTKAPKPTSTKPKTTPRVRKPTTTPRXTMTSTPELN 859
QY 901 TTEKTTESKIATTTTQVSTTTQDTPPKITTLTKTTILAPKVTTTKIITTEIMNKPEE 960
Db 860 TTEKTTESKIATTTTQVSTTTQDTPPKITTLTKTTILAPKVTTTKIITTEIMNKPEE 919
QY 961 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPTTTPRXTMTSTPELN 1020
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPTTTPRXTMTSTPELN 979
QY 1021 PTSRIAEMLQTTTRPNQTPNSKLEVNPKSEDAGAGETPHMLLRHVFMPVETPDM 1080
Db 980 PTSRIAEMLQTTTRPNQTPNSKLEVNPKSEDAGAGETPHMLLRHVFMPVETPDM 1039
QY 1081 YLPRVFNCGIILNPMLEDETNIENCGKPDVGLTTLNGLTVAPRGHYFWMLSPFSPSPAR 1140
Db 1040 YLPRVFNCGIILNPMLEDETNIENCGKPDVGLTTLNGLTVAPRGHYFWMLSPFSPSPAR 1099
QY 1141 RITEVWGIPIPIDTVFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTQI 1200
Db 1100 RITEVWGIPIPIDTVFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTQI 1159
QY 1201 VAALSTAKYNWPEVSYFFKRGGSICQYIYKQEPVQKCPGRPALNYPYGBMTQVRRRR 1260
Db 1160 VAALSTAKYNWPEVSYFFKRGGSICQYIYKQEPVQKCPGRPALNYPYGBMTQVRRRR 1219
QY 1261 FERAIQPSQTHIRIQYSFARLAYQDKGVLHNEVKVSIILWRLPLNVVTSISAIPLNIRKPD 1320
Db 1220 FERAIQPSQTHIRIQYSFARLAYQDKGVLHNEVKVSIILWRLPLNVVTSISAIPLNIRKPD 1279
QY 1321 GYDYVAFSKQYQYINIDVPSRTARAITTRSGQTLISKWYNCP 1361
Db 1280 GYDYVAFSKQYQYINIDVPSRTARAITTRSGQTLISKWYNCP 1320

RESULT 8

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO:

60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-60

Query Match

Best Local Similarity

Matches 1320; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Score 7058.9; DB 4; Length 1320;

Pred. No. 5.8e-194;

QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCHYMCCPDF 60

Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQ----- 25

QY 61 KRVTAEISCKRCRCFESFERGECDCDAOCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 120

Db 26 -----ELSCRCRCFESFERGECDCDAOCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 79

QY 121 PPSGASQTIKSTTKSPKPPNKKTKVIESEIEITEVDKNKNRTKKKPTPKPPVVDKAG 180

Db 80 PPSGASQTIKSTTKSPKPPNKKTKVIESEIEITEVDKNKNRTKKKPTPKPPVVDKAG 139

QY 181 SGLDNGDFKVTTPDTSTTQHNNKSVSPKITTAKPINPRESLPPNSDTSKETSITVNNKETT 240

Db 140 SGLDNGDFKVTTPDSTTTOHNVSTSPKLTAKPINPESLSPNSDTSKETSITVKNKETT 199
QY 241 VETKEITTTNKOTSDGKEKTTSAKETSIEKTSKADLAPTAKVIAKPTPKAETTTKGP 300
Db 200 VETKEITTTNKOTSDGKEKTTSAKETSIEKTSKADLAPTAKVIAKPTPKAETTTKGP 259
QY 301 LTTPEPTTTPKEPASTTTPKEPTTTIKSAPTTKEPAPTTPKAPTTTKEPAPTTPKE 360
Db 260 LTTPEPTTTPKEPASTTTPKEPTTTIKSAPTTKEPAPTTPKAPTTTKEPAPTTPKE 319
QY 361 PAPTTPKEPAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKEPTTT 420
Db 320 PAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPTTT 379
QY 421 PKBPAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 480
Db 380 PKBPAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 439
QY 481 PTTTKSAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 540
Db 440 PTTTKSAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 499
QY 541 PAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKEPAPT 600
Db 500 PAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKEPAPT 559
QY 601 TPELAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 660
Db 560 TPELAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 619
QY 661 TAPTTPKGAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPK 720
Db 620 TAPTTPKGAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTPK 679
QY 721 EPAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 780
Db 680 EPAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 739
QY 781 KETAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 840
Db 740 KETAPTTPKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 799
QY 841 PTPKALENSKPEPGVPTTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 900
Db 800 PTPKALENSKPEPGVPTTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTKEPAPTTPKAPTTTP 859
QY 901 TTEKTTESKITTATTTQVSTTTTQDTPPKITTLKTTTAPKVTITTKITTTTEIMNKP 960
Db 860 TTEKTTESKITTATTTQVSTTTTQDTPPKITTLKTTTAPKVTITTKITTTTEIMNKP 919
QY 961 TAKPKDRATNSKATTPKAPKPTKAPKPTSTKPKTTPKPKTTPKPKTTPKPKTTPKPKTTPKPKTTP 1020
Db 920 TAKPKDRATNSKATTPKAPKPTKAPKPTSTKPKTTPKPKTTPKPKTTPKPKTTPKPKTTPKPKTTP 979
QY 1021 PTSRIAEMLOTTTRPNQTPNSKLVEVPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 980 PTSRIAEMLOTTTRPNQTPNSKLVEVPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1081 YLPRVNOGIIINPMLSDENIINGKPVGDGLTLRNGTLVAFRGHYFWMLSPPSPSPAR 1140
Db 1040 YLPRVNOGIIINPMLSDENIINGKPVGDGLTLRNGTLVAFRGHYFWMLSPPSPSPAR 1099
QY 1141 RITEVWGIPIPIDTVFTRCNCEGTTFKDSQVWRFTNDIKDAGYKPKPIFKGGLGTQI 1200
Db 1100 RITEVWGIPIPIDTVFTRCNCEGTTFKDSQVWRFTNDIKDAGYKPKPIFKGGLGTQI 1159
QY 1201 VAALSTAKYKNWPESVYFFKRGSIQOYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRR 1260
Db 1160 VAALSTAKYKNWPESVYFFKRGSIQOYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRR 1219
QY 1261 FERAIGPSQTHTRIOYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1320

Db 1220 FERAIGPSQTHTRIOYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1279
QY 1321 GVDYVAFSKDOYINIDVFSRTARAITTRSGQTLSKVWYNCP 1361
Db 1280 GVDYVAFSKDOYINIDVFSRTARAITTRSGQTLSKVWYNCP 1320
RESULT 9
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-42
Query Match 96.3%; Score 7049; DB 4; Length 1311;
Best Local Similarity 96.3%; Pred. No. 1.le-191;
Matches 1311; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVYSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVYSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGECDCDAOCKYDKCCPDYSPCAEVHNPTSPSSKAP 120
Db 61 KRVTCTAELCKGRCFESFERGECDCDAOCKYDKCCPDYSPCAEVHNPTSPSSKAP 105

QY 121 PPSGASQTIKSTTKRSKPPNKKTKKXVISEEELITEVKNKKNKTKKKPTKPPVDEAG 180
Db 106 -----EVKDNKKNKTKKKPTKPPVDEAG 130
QY 181 SGLDNGDFKVTTPDTSTTQHNKYSTSKPIITAKBINRPSLPPNSDTSKETSLTVNKETT 240
Db 131 SGLDNGDFKVTTPDTSTTQHNKYSTSKPIITAKBINRPSLPPNSDTSKETSLTVNKETT 190
QY 241 VETKETTITNKQTSIDGKEKTSIAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 300
Db 191 VETKETTITNKQTSIDGKEKTSIAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 250
QY 301 LTTPEKEPTTPKEPASTTPKEPTTTIKSAPTTKGPAPTTTKSAPTTKSAPTTPKEPAPTTTKE 360
Db 251 LTTPEKEPTTPKEPASTTPKEPTTTIKSAPTTKGPAPTTTKSAPTTKSAPTTPKEPAPTTTKE 310
QY 361 PAPTTPKEPAPTTTKAPATTTKSAPTTKGPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTT 420
Db 311 PAPTTPKEPAPTTTKAPATTTKSAPTTKGPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTT 370
QY 421 PKEPAPTTKEPAPTTPKAPATAPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 480
Db 371 PKEPAPTTKEPAPTTPKAPATAPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 430
QY 481 PTTTKSAPTTTKAPATTTKSAPTTKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 540
Db 431 PTTTKSAPTTTKAPATTTKSAPTTKGPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 490
QY 541 PAPTTPKEPAPTTTKKAPATAPKAPATTPKEPAPTTTPKLTPTTPEKLAPTTPEKPAPT 600
Db 491 PAPTTPKEPAPTTTKKAPATAPKAPATTPKEPAPTTTPKLTPTTPEKLAPTTPEKPAPT 550
QY 601 TPELAPTTPEEPTTPEEAPPTPKAAANWPKAPATTPKAPATTPKAPATTPKAPATTPKE 660
Db 551 TPELAPTTPEEPTTPEEAPPTPKAAANWPKAPATTPKAPATTPKAPATTPKAPATTPKE 610
QY 661 TAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKETSTTSKAPATTPKGTAPTTPK 720
Db 611 TAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKETSTTSKAPATTPKGTAPTTPK 670
QY 721 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTP 780
Db 671 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTP 730
QY 781 KETAPTTKEPAPTTPKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840
Db 731 KETAPTTKEPAPTTPKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPESTPELSAE 790
QY 841 PTEKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 900
Db 791 PTEKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 850
QY 901 TTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEE 960
Db 851 TTEKTTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEE 910
QY 961 TAKPKDRATNSKATTPKOKPKAPKKTSTIKPKMTMRVRPKTTPPKMTSTMPBLN 1020
Db 911 TAKPKDRATNSKATTPKOKPKAPKKTSTIKPKMTMRVRPKTTPPKMTSTMPBLN 970
QY 1021 PTSRIAEAMLQTTTRNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPEVTPDMD 1080
Db 971 PTSRIAEAMLQTTTRNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPEVTPDMD 1030
QY 1081 YLPRVNPQGIITINPMLSDETNICNGKPVVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPAR 1140
Db 1031 YLPRVNPQGIITINPMLSDETNICNGKPVVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPAR 1090
QY 1141 RITEVWGIPSPIDTVTRCNCEKTFPFKDSQVWRFTNDIKDAGVPKPIFKFGGLTGQI 1200
Db 1091 RITEVWGIPSPIDTVTRCNCEKTFPFKDSQVWRFTNDIKDAGVPKPIFKFGGLTGQI 1150
QY 1201 VAALSTAKYNWPESVYFFKRGGSIOQYIYKQBPVKCGRPALNYPVYGMTQVRRRR 1260

Db 1151 VAALSTAKYNWPESVYFFKRGGSIOQYIYKQBPVKCGRPALNYPVYGMTQVRRRR 1210
QY 1261 FERAIGSPQTHTRIQYSPARLAYODKGVILHNEVKVKSILWRGLPNVVTSAISLENIRKPD 1320
Db 1211 FERAIGSPQTHTRIQYSPARLAYODKGVILHNEVKVKSILWRGLPNVVTSAISLENIRKPD 1270
QY 1321 GYDYAFSKQOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1361
Db 1271 GYDYAFSKQOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1311
RESULT 10
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseiz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52
Query Match 96.2%; Score 7044.6; DB 4; Length 1363;
Best Local Similarity 94.0%; Pred. No. 1.5e-191;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;
QY 1 MAWKTLPIYLLLLLSVFIQVSSQDLSSCAGCGEGYSRDATCNCDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFIQVSSQ----- 25

QY 61 KRVTAEISCKGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCKGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----156
DB 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEEHSVSENQESSSSSSSSSSTIw 139
QY 157 -----VKONKNRTKKKPTPKPPVVDAGSGLDNGDFKVTTPDTST 197
DB 140 KIKSSKNSAANRELQKKLVKONKNRTKKKPTPKPPVVDAGSGLDNGDFKVTTPDTST 199
QY 198 TQHNKYSTSKITITAKPINRPSLPNSDTSKETSILVNKETTVEKETTNNKQSTDGD 257
DB 200 TQHNKYSTSKITITAKPINRPSLPNSDTSKETSILVNKETTVEKETTNNKQSTDGD 259
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 317
DB 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 319
QY 318 TTPKEPTPTTIKSAPTTPKEPAPTTTTSKAPTTTPKEPAPTTTTPKEPAPTTTTPKEP 377
DB 320 TTPKEPTPTTIKSAPTTPKEPAPTTTTSKAPTTTPKEPAPTTTTPKEPAPTTTTPKEP 379
QY 378 APPTTTSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTTPKEPAPTTTPK 437
DB 380 APPTTTSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTTPKEPAPTTTPK 439
QY 438 EPAPTAPKKAPAPTTTPKEPAPTTTPKPSPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 497
DB 440 EPAPTAPKKAPAPTTTPKEPAPTTTPKPSPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 499
QY 498 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKEPAPTTTTPKEPAPTT 557
DB 500 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKEPAPTTTTPKEPAPTT 559
QY 558 APAPTAPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKEPAPTTTTP 617
DB 560 APAPTAPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKEPAPTTTTP 619
QY 618 PEPAPPTTPKAAANPTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTT 677
DB 620 PEPAPPTTPKAAANPTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTT 679
QY 678 APPTPKAPAPAPAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 737
DB 680 APPTPKAPAPAPAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 739
QY 738 TAPPTILKEPAPPTTPKAPAPAPAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 797
DB 740 TAPPTILKEPAPPTTPKAPAPAPAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 799
QY 798 KPAPTTPETPPPTTSEVSTPTTKBPTTIHKSPDESTPELSABPTPKALENSPKAPVPT 857
DB 800 KPAPTTPETPPPTTSEVSTPTTKBPTTIHKSPDESTPELSABPTPKALENSPKAPVPT 859
QY 858 TKTPAATKPMPTTAKDITERRDLRTTPETTTAAPKMTKATATTTEKTTESKITATTQV 917
DB 860 TKTPAATKPMPTTAKDITERRDLRTTPETTTAAPKMTKATATTTEKTTESKITATTQV 919
QY 918 TSTTTQDTPFKITLTKITTLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK 977
DB 920 TSTTTQDTPFKITLTKITTLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK 979
QY 978 PQKPTKAPKKPTSTKPKMVRVKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1037
DB 980 PQKPTKAPKKPTSTKPKMVRVKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1039
QY 1038 QTNNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1097
DB 1040 QTNNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1099

QY 1098 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITTEVWGIPSPIDTVP 1157
DB 1100 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITTEVWGIPSPIDTVP 1159
QY 1158 RCNCEGXTFFFKDSQYWRFTNDIKDAGYKPIPKGFGGLTGOIIVAALSTAKYKWPESVY 1217
DB 1160 RCNCEGXTFFFKDSQYWRFTNDIKDAGYKPIPKGFGGLTGOIIVAALSTAKYKWPESVY 1219
QY 1218 FFKRGSGSIQYIYKQEPVQKCGRRPALNVPYVGMTQVRRRRFERAIGPSQHTIRIQY 1277
DB 1220 FFKRGSGSIQYIYKQEPVQKCGRRPALNVPYVGMTQVRRRRFERAIGPSQHTIRIQY 1279
QY 1278 SPARLAYODKGVHNEVKVLSILMRGLPNVVTSAISLEINIRKPDGYDYAFSKDQYNNIDV 1337
DB 1280 SPARLAYODKGVHNEVKVLSILMRGLPNVVTSAISLEINIRKPDGYDYAFSKDQYNNIDV 1339
QY 1338 PSRTARAITTRSGQTLSSKWNVNC 1361
DB 1340 PSRTARAITTRSGQTLSSKWNVNC 1363
RESULT 11
US-10-164-595-58
; Sequence 58, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 58
; LENGTH: 1320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-58
Query Match 96.2%; Score 7043.9; DB 4; Length 1320;
Best Local Similarity 96.8%; Pred. No. 1.5e-191;
Matches 1317; Conservative 0; Mismatches 3; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVVIQQVSSQDLSSCAGCGEGYSRDATCNCYNCQHMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVVIQQVSSQ-----25
QY 61 KRVTAEISCKGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCKGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEVKNKKNRTKKKPTPKPPVVDAG 180
DB 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEVKNKKNRTKKKPTPKPPVVDAG 139
QY 181 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETT 240
DB 140 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETT 199
QY 241 VETKETTTNNKOTSDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 300
DB 200 VETKETTTNNKOTSDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 259
QY 301 LITPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTTTPKAPAPTTTTPKE 360
DB 260 LITPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTTTPKAPAPTTTTPKE 319
QY 361 PAPTTPKEPAPTTTPKEPAPTTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKEPTPTT 420
DB 320 PAPTTPKEPAPTTTPKEPAPTTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKEPTPTT 379
QY 421 PKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKEPTPTTKEPA 480

QY 301 LITPKPTPTTKEPASTTKEPTTITKSAPTTPKEPATTTKSAPTTKEBPATTTKE 360
DB 253 LITPKPTPTTKEPASTTKEPTTITKSAPTTPKEPATTTKSAPTTKEBPATTTKE 312
QY 361 PAPTTPKEPATTTTKEPATTTKSAPTTPKEPATTTKKAAPTTPKEPATTTKEPTPTT 420
DB 313 PAPTTPKEPATTTTKEPATTTKSAPTTPKEPATTTKKAAPTTPKEPATTTKEPTPTT 372
QY 421 PKEPATTTKEPATTTTKEPATTTKKAAPTTPKEPATTTTKEPATTTTKEPSPTTKEPA 480
DB 373 PKEPATTTKEPATTTTKEPATTTKKAAPTTPKEPATTTTKEPATTTTKEPSPTTKEPA 432
QY 481 PTTTKEPATTTTKEPATTTKSAPTTPKEPSPTTKEPATTTTKEPATTTTKEPATTTTKE 540
DB 433 PTTTKEPATTTTKEPATTTKSAPTTPKEPSPTTKEPATTTTKEPATTTTKEPATTTTKE 492
QY 541 PAPTTPKEPATTTTKEPATTTKKAAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTT 600
DB 493 PAPTTPKEPATTTTKEPATTTKKAAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTT 552
QY 601 TPELATTTPEPTPTTPEEPATTPKAAAPTTPKEPATTTTKEPATTTTKEPATTTTKE 660
DB 553 TPELATTTPEPTPTTPEEPATTPKAAAPTTPKEPATTTTKEPATTTTKEPATTTTKE 612
QY 661 TAPTTPKGATTTTKEPATTTKKAAPTTPKELAPTTTKEPATTTTKEPATTTTKEPATTT 720
DB 613 TAPTTPKGATTTTKEPATTTKKAAPTTPKELAPTTTKEPATTTTKEPATTTTKEPATTT 672
QY 721 EPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 780
DB 673 EPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 732
QY 781 KETAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 840
DB 733 KETAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 792
QY 841 PTKALENSPKEPGVPTTTPAAKPEMTTAKOKTTERDLRTTTPETTTAAPTTPKAT 900
DB 793 PTKALENSPKEPGVPTTTPAAKPEMTTAKOKTTERDLRTTTPETTTAAPTTPKAT 852
QY 901 TTEKTTESKIATTTTQVTSITTTQDTTPFKITLTKITTLAPKVTTTKITTTTEIMNKEE 960
DB 853 TTEKTTESKIATTTTQVTSITTTQDTTPFKITLTKITTLAPKVTTTKITTTTEIMNKEE 912
QY 961 TAKPKDRATNKATTPKQKPTKAPKKTSTKPKTTPRVRKPTTTPRKTSTMPBLN 1020
DB 913 TAKPKDRATNKATTPKQKPTKAPKKTSTKPKTTPRVRKPTTTPRKTSTMPBLN 972
QY 1021 PTSRIAEAMLQTTTPRNPOTNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1080
DB 973 PTSRIAEAMLQTTTPRNPOTNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1032
QY 1081 YLPRVNPQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1140
DB 1033 YLPRVNPQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1092
QY 1141 RITEVWGIPSPIDTFTTRCNCEGKTFPFKDSQYWFRTNDIKDAGYKPKIFKFGGLTGOI 1200
DB 1093 RITEVWGIPSPIDTFTTRCNCEGKTFPFKDSQYWFRTNDIKDAGYKPKIFKFGGLTGOI 1152
QY 1201 VAALSTAKYKNWPFESVYFPRGSGSQYIYKQEPVQKCPGRPALNYPVYGEMTOVRRRR 1260
DB 1153 VAALSTAKYKNWPFESVYFPRGSGSQYIYKQEPVQKCPGRPALNYPVYGEMTOVRRRR 1212
QY 1261 FERAIGPSQTHIRIQYSPARLAYQDKVGLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1320
DB 1213 FERAIGPSQTHIRIQYSPARLAYQDKVGLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1272
QY 1321 GHDYAFSKDQYINIDVPRTARAITTRSGQTLISKWYNCP 1361
DB 1273 GHDYAFSKDQYINIDVPRTARAITTRSGQTLISKWYNCP 1313

RESULT 14
US-07-757-022B-44
; Sequence 44, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; LENGTH: 1270 amino acids
; SEQUENCE CHARACTERISTICS:
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-44
Query Match 92.7%; Score 6784.9; DB 4; Length 1270;
Best Local Similarity 93.3%; Pred. No. 3.1e-184;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLILLLSVFVIQVSSQDLSSCAGCGEGYSDATCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLILLLSVFVIQVSSQ----- 25
QY 61 KRVCCTAELSCGRCPESFERGREGCDCAQCKKYDKCCPDYEFCAEHNPTSPSSKKAP 120
DB 26 -----ELSCGRCPESFERGREGCDCAQCKKYDKCCPDYEFSCA----- 64
QY 121 PEGSAGQTIKSTTKSPKPPNKKTKVIESEITEVKNKNRTPKPTPKPPVVDKAG 180
DB 65 -----EVKNKNRTPKPTPKPPVVDKAG 89
QY 181 SGLDNGDKVTTPTDSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 240
DB 90 SGLDNGDKVTTPTDSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 149

QY 241 VETKETTITNKQTSIDGKEKTTISAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKGPA 300
Db 150 VETKETTITNKQTSIDGKEKTTISAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKGPA 209
QY 301 LITPKBPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 360
Db 210 LITPKBPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 269
QY 361 PAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 420
Db 270 PAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 329
QY 421 PKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKE 480
Db 330 PKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKEPAPTITKE 389
QY 481 PTTTKSAPTITKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 540
Db 390 PTTTKSAPTITKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITKE 449
QY 541 PAPTTPKEPAPTITTKKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPAT 600
Db 450 PAPTTPKEPAPTITTKKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPATPKAPAT 509
QY 601 TPEELAPTTPEBPTTTPKEPAPTITPKAAAPTTPKEPAPTITPKAAAPTTPKEPAPTITPKAAAPTTPKE 660
Db 510 TPEELAPTTPEBPTTTPKEPAPTITPKAAAPTTPKEPAPTITPKAAAPTTPKEPAPTITPKAAAPTTPKE 569
QY 661 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTPK 720
Db 570 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTPK 629
QY 721 BPAPTTPKEPAPTITPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTPK 780
Db 630 BPAPTTPKEPAPTITPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTPK 689
QY 781 KETAPTTPKEPAPTITPKKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 840
Db 690 KETAPTTPKEPAPTITPKKAPATTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 749
QY 841 PTPKALNSPKBPGVPTTKTAAPKPEMTTAKOKTTERDLRTTPTTAAAPKMTKETAT 900
Db 750 PTPKALNSPKBPGVPTTKTAAPKPEMTTAKOKTTERDLRTTPTTAAAPKMTKETAT 809
QY 901 TTEKTESKITATTQTSTTTQDTPPKITTLKTTTLAPKVTITTKKITTTEIMNKPEE 960
Db 810 TTEKTESKITATTQTSTTTQDTPPKITTLKTTTLAPKVTITTKKITTTEIMNKPEE 869
QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMRVRKPKTTPTRKWTSTMPBLN 1020
Db 870 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMRVRKPKTTPTRKWTSTMPBLN 929
QY 1021 PTSIAEAMLOTTTRPNQTPNSKLVVNPKSEDGAGGETPHMLLRPHVFMPEVTPDMD 1080
Db 930 PTSIAEAMLOTTTRPNQTPNSKLVVNPKSEDGAGGETPHMLLRPHVFMPEVTPDMD 989
QY 1081 YLPRVFNQGIILNPMLSDETNICNGKPDGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1140
Db 990 YLPRVFNQGIILNPMLSDETNICNGKPDGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1049
QY 1141 RITEVWGIPIPIDVTRFRCNCEGKTFKFDKSYWRFTNDIKDAGYKPKPIFKGFGGLTGQI 1200
Db 1050 RITEVWGIPIPIDVTRFRCNCEGKTFKFDKSYWRFTNDIKDAGYKPKPIFKGFGGLTGQI 1109
QY 1201 VAALSTAKYKNWPSVVFVKRGGISIQYIYKQEPVQKPCGRRPALNYPVYGMTQVRRRR 1260
Db 1110 VAALSTAKYKNWPSVVFVKRGGISIQYIYKQEPVQKPCGRRPALNYPVYGMTQVRRRR 1169
QY 1261 FERAIGPSQTHIRIQSPARLAYODKGLVHNEVKVSVILMRGLNVVTSALSINIRKPD 1320
Db 1170 FERAIGPSQTHIRIQSPARLAYODKGLVHNEVKVSVILMRGLNVVTSALSINIRKPD 1229

QY 1321 GYDYAFSKDQYINIDVPSRTARAITTRSGQTLISKVWYNCP 1361
Db 1230 GYDYAFSKDQYINIDVPSRTARAITTRSGQTLISKVWYNCP 1270

RESULT 15

US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-104

Query Match 80.1%; Score 5867.7; DB 4; Length 1140;
Best Local Similarity 96.2%; Pred. No. 2.5e-158;
Matches 1097; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNQCHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNQCHYMECCPDF 60
QY 61 KRVTABLSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSSKAP 120
Db 61 KRVTABLSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSSKAP 120
QY 121 PPSSASQTIKSTTKRSPKPNKKTKVIESEITE----- 156
Db 121 PPSSASQTIKSTTKRSPKPNKKTKVIESEITE----- 156

Db 121 PPGASQTIKSTTRSPKPPNKKTKKVISEBITEHSHSVSENQESSSSSSSSSSSSSTIW 180
QY 157 -----VKDNKNRTKKKPTKPPVVDVAGSGLONGDFKVITPDTST 197
Db 181 KIKSSKNSAANRELQKKLVKDNKNKNRTKKKPTKPPVVDVAGSGLONGDFKVITPDTST 240
QY 198 TOHNKVSTSPKITTAKPILNPPSLPPNSDTSKETSILVNKETTIVETKETTITNKQTSIDG 257
Db 241 TOHNKVSTSPKITTAKPILNPPSLPPNSDTSKETSILVNKETTIVETKETTITNKQTSIDG 300
QY 258 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTKPAETTTKGPAITTKPEPTTKPEPTTKPEAS 317
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTKPAETTTKGPAITTKPEPTTKPEPTTKPEAS 360
QY 318 TTPKEPTTTIKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 377
Db 361 TTPKEPTTTIKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY 378 APTTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 437
Db 421 APTTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 480
QY 438 EPAPTAPKKPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 497
Db 481 EPAPTAPKKPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 540
QY 498 TTKSAPTTPKBPSPPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 557
Db 541 TTKSAPTTPKBPSPPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 600
QY 558 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 617
Db 601 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 660
QY 618 PREPAPTTPKAAAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 677
Db 661 PREPAPTTPKAAAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 720
QY 678 APTTPKBPAPKELAPTTTKBPSTTSKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 737
Db 721 APTTPKBPAPKELAPTTTKBPSTTSKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 780
QY 738 TAPTTILKBPATTPKBPAPKELAPTTTKGPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 797
Db 781 TAPTTILKBPATTPKBPAPKELAPTTTKGPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 840
QY 798 KDPATTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKBPGVPT 857
Db 841 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKBPGVPT 900
QY 858 TKTTPAATKPEMTTAKDKTTERDLRTPPETTTAAAPKOTKETATTTKETTISKITATTQV 917
Db 901 TKTTPAATKPEMTTAKDKTTERDLRTPPETTTAAAPKOTKETATTTKETTISKITATTQV 960
QY 918 TSTTTQDTPPKITTLTKTTLLAPKVTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 977
Db 961 TSTTTQDTPPKITTLTKTTLLAPKVTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 978 FOKPTKAPKKPTSTKKPKTTPVRVRKPTTTPPKMTSTMPELNPTSRIAEAMLQTTTREN 1037
Db 1021 FOKPTKAPKKPTSTKKPKTTPVRVRKPTTTPPKMTSTMPELNPTSRIAEAMLQTTTREN 1080
QY 1038 QTPNSKLIVEVNPSEBAGGEGTTPHMLLRPHVFMPEVTPDMDYLPVFNQGLIINPMLS 1097
Db 1081 QTPNSKLIVEVNPSEBAGGEGTTPHMLLRPHVFMPEVTPDMDYLPVFNQGLIINPMLS 1140

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 138.466 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQLSKVWVNC 1361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7323	100.0	1361	13	US-10-124-557-40
2	7308.7	99.8	1404	9	US-09-802-207-30
3	7308.7	99.8	1404	11	US-09-897-188-1
4	7308.7	99.8	1404	13	US-10-124-557-2
5	7308.7	99.8	1404	13	US-10-124-557-62
6	7084.3	96.7	1354	13	US-10-124-557-48
7	7058.9	96.4	1320	13	US-10-124-557-46
8	7058.9	96.4	1320	13	US-10-124-557-60
9	7049	96.3	1311	13	US-10-124-557-42
10	7044.6	96.2	1363	13	US-10-124-557-52
11	6836.3	93.4	1314	13	US-10-124-557-50
12	6820.2	93.1	1313	13	US-10-124-557-142
13	6784.9	92.7	1270	13	US-10-124-557-44
14	5867.7	80.1	1140	13	US-10-124-557-104

15	5485.7	74.9	1038	13	US-10-124-557-74	Sequence 74, Appl
16	5472	74.7	1022	13	US-10-124-557-84	Sequence 84, Appl
17	5379.2	73.5	1049	13	US-10-124-557-58	Sequence 58, Appl
18	5011	68.4	941	13	US-10-124-557-14	Sequence 14, Appl
19	3944.5	53.9	792	9	US-09-802-207-27	Sequence 27, Appl
20	2850.9	38.9	538	14	US-10-038-694-3	Sequence 3, Appl
21	2246.2	30.7	463	13	US-10-124-557-54	Sequence 54, Appl
22	1991.2	27.2	423	13	US-10-124-557-66	Sequence 66, Appl
23	1982.1	27.1	422	13	US-10-124-557-68	Sequence 68, Appl
24	1723.1	23.5	372	13	US-10-124-557-64	Sequence 64, Appl
25	1714.7	23.4	401	9	US-09-802-207-29	Sequence 4, Appl
26	1566.9	21.4	292	16	US-10-468-910-4	Sequence 1068, Appl
27	1413.6	19.3	5179	9	US-09-922-217-1068	Sequence 1068, Ap
28	1413.6	19.3	5179	9	US-09-833-263-1068	Sequence 1068, Ap
29	1413.6	19.3	5179	13	US-10-025-380-1068	Sequence 1068, Ap
30	1413.6	19.3	5179	16	US-10-734-564-121	Sequence 121, Appl
31	1295.5	17.7	296	13	US-10-124-557-70	Sequence 70, Appl
32	1214.6	16.6	237	13	US-10-124-557-72	Sequence 72, Appl
33	1143	15.6	207	13	US-10-124-557-116	Sequence 116, Appl
34	1143	15.6	207	13	US-10-124-557-136	Sequence 136, Appl
35	1025.9	14.0	185	16	US-10-468-910-2	Sequence 2, Appl
36	1010	13.8	188	14	US-10-038-694-2	Sequence 2, Appl
37	992.5	13.6	1460	14	US-10-295-027-428	Sequence 428, Appl
38	983.3	13.4	3507	14	US-10-369-493-5784	Sequence 5784, Ap
39	971.4	13.3	6642	14	US-10-369-493-5013	Sequence 5013, Ap
40	971	13.3	1367	9	US-09-801-368-108	Sequence 108, Appl
41	970.7	13.3	220	13	US-10-124-557-96	Sequence 96, Appl
42	957.9	13.1	19723	15	US-10-084-846A-5	Sequence 5, Appl
43	956	13.1	231	13	US-10-124-557-30	Sequence 30, Appl
44	953.9	13.0	1325	9	US-09-864-761-35612	Sequence 35612, A
45	922.8	12.6	5935	14	US-10-243-243A-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 100.0%; Score 7323; DB 13; Length 1361;
Best Local Similarity 100.0%; Pred. No. 1.7e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KRVCIAELCKGRGCFESFERGECDDCAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Db 61 KRVCIAELCKGRGCFESFERGECDDCAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTRKSPKPPNKKTKVIESEITEVKDNKKNRTKKKPTKPPVDEAG 180
Db 121 PPSGASQTIKSTTRKSPKPPNKKTKVIESEITEVKDNKKNRTKKKPTKPPVDEAG 180
QY 181 SGLDNGDFKVTPTDSTTHNKNVSTSPKITTAKINPRPSLPNSDTSKETSIVNKEIT 240
Db 181 SGLDNGDFKVTPTDSTTHNKNVSTSPKITTAKINPRPSLPNSDTSKETSIVNKEIT 240
QY 241 VETKETTTNQTSDGKEKTTSAKETQSTKTSKOLAPTSKVLAKPTPKAETTTKGA 300
Db 241 VETKETTTNQTSDGKEKTTSAKETQSTKTSKOLAPTSKVLAKPTPKAETTTKGA 300
QY 301 LTTPKETPTTPKBPASTTPKEPTTTIKSAPTTKBPAPTTPKBPAPTTPKE 360
Db 301 LTTPKETPTTPKBPASTTPKEPTTTIKSAPTTKBPAPTTPKBPAPTTPKE 360
QY 361 PAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 420
Db 361 PAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 420
QY 421 PKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 480
Db 421 PKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 480
QY 481 PTTTKSAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 540
Db 481 PTTTKSAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 540
QY 541 PAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 600
Db 541 PAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 600
QY 601 TPELAPTTPKEPTTPPEBPAPTTPKAAAPNTKPEPAPTTPKBPAPTTPKBPAPT 660
Db 601 TPELAPTTPKEPTTPPEBPAPTTPKAAAPNTKPEPAPTTPKBPAPTTPKBPAPT 660
QY 661 TAPPTPKGAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
Db 661 TAPPTPKGAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
QY 721 EPAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 780
Db 721 EPAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 780
QY 781 KETAPTTPKBPAPTTPKBPAPTTPETPPPTTSEVSTPTTTKPTTIHKSPDESP 840

Db 781 KETAPTTPKBPAPTTPKBPAPTTPETPPPTTSEVSTPTTTKPTTIHKSPDESP 840
QY 841 PTPKALENSPKPEPGVPTTKTAAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
Db 841 PTPKALENSPKPEPGVPTTKTAAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
QY 901 TTEXTTESKITATTTQVTSITTTQDTPPKITTTTLTKTLAPKVTTKTITTTTEIMNKPEE 960
Db 901 TTEXTTESKITATTTQVTSITTTQDTPPKITTTTLTKTLAPKVTTKTITTTTEIMNKPEE 960
QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTRKWTSTMPSELN 1020
Db 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTRKWTSTMPSELN 1020
QY 1021 PTSRIAEMLQTTTRPNQTPNSKLVEVNPXSSEDAGGAETPHMLLRPHVFMPEVTPDMD 1080
Db 1021 PTSRIAEMLQTTTRPNQTPNSKLVEVNPXSSEDAGGAETPHMLLRPHVFMPEVTPDMD 1080
QY 1081 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1140
Db 1081 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1140
QY 1141 RITEVWGIPIPDVFTFRCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQOI 1200
Db 1141 RITEVWGIPIPDVFTFRCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQOI 1200
QY 1201 VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVQKCPGRRPALNPVYVGMTOVRRRR 1260
Db 1201 VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVQKCPGRRPALNPVYVGMTOVRRRR 1260
QY 1261 FERAIGPSQTHIRIOYSPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLENIRKPD 1320
Db 1261 FERAIGPSQTHIRIOYSPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLENIRKPD 1320
QY 1321 GYDYAFSKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1321 GYDYAFSKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361

RESULT 2

US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US20020086824A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match 99.8%; Score 7308.7; DB 9; Length 1404;
Best Local Similarity 96.9%; Pred. No. 3.9e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
QY 157 -----VKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
DB 181 KIKSSKNNSAANRELQKKLVKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 198 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 257
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 300
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
QY 318 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 377
DB 361 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 420
QY 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
QY 157 -----VKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
DB 181 KIKSSKNNSAANRELQKKLVKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 198 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 257
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 300
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
QY 318 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 377
DB 361 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 420

DB 1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSIDTFT 1200
QY 1158 RCNCEGKTFEFDKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTGQIIVAALSTAKYKNWPSVY 1217
DB 1201 RCNCEGKTFEFDKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTGQIIVAALSTAKYKNWPSVY 1260
QY 1218 FFKRGSGIOQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRRPERAIGPSQTHIRIQY 1277
DB 1261 FFKRGSGIOQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRRPERAIGPSQTHIRIQY 1320
QY 1278 SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPDGYAFSKQOYNNIDV 1337
DB 1321 SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPDGYAFSKQOYNNIDV 1380
QY 1338 PSRTARAITTRSGQTLSKWKYNCP 1361
DB 1381 PSRTARAITTRSGQTLSKWKYNCP 1404
RESULT 3
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1
Query Match 99.8%; Score 7308.7; DB 11; Length 1404;
Best Local Similarity 96.8%; Pred. No. 3.9e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCBGYSRDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCBGYSRDATCNCYNCOHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
DB 121 PPSGASQTIKSTTKRSKPPNKKTKVIESEITE----- 156
QY 157 -----VKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
DB 181 KIKSSKNNSAANRELQKKLVKDNKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 198 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 257
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKNQSTGDG 300
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
QY 318 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 377
DB 361 TTPKEPTPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTKBP 420

QY 258 KEKTTSAKETOSIEKTSADLAPTSSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 317
Db 301 KEKTTSAKETOSIEKTSADLAPTSSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 360
QY 318 TTPKEPTTTIKSAPTTPKEPATTTKGAPTTPKEPATTTKBPATTTKBPATTTKBP 377
Db 361 TTPKEPTTTIKSAPTTPKEPATTTKGAPTTPKEPATTTKBPATTTKBPATTTKBP 420
QY 378 APTTTKSAPTTPKEPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 437
Db 421 APTTTKSAPTTPKEPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 480
QY 438 EPAPTAPKKAPTTPKEPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 497
Db 481 EPAPTAPKKAPTTPKEPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 540
QY 498 TTKSAPTTPKEPATTTKBPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 557
Db 541 TTKSAPTTPKEPATTTKBPATTTPKKPAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 600
QY 558 APTAPKEPATTPKETAPTTPKELTPTTPKELAPTTPKEPATTTPEELAPTTPKEPATTT 617
Db 601 APTAPKEPATTPKETAPTTPKELTPTTPKELAPTTPKEPATTTPEELAPTTPKEPATTT 660
QY 618 PEPAPTTPKAAAPTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTTP 677
Db 661 PEPAPTTPKAAAPTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTTP 720
QY 678 APTPKKPAKELAPTTPKEPATTTSDKAPTTPKGAAPTTPKEPATTTKBPATTTKBPATTTKBP 737
Db 721 APTPKKPAKELAPTTPKEPATTTSDKAPTTPKGAAPTTPKEPATTTKBPATTTKBPATTTKBP 780
QY 738 TAPTTLKEPATTPPKKPAKELAPTTPKGTSTSDKAPTTPKETAAPTTPKEPATTTKBPATTTKBP 797
Db 781 TAPTTLKEPATTPPKKPAKELAPTTPKGTSTSDKAPTTPKETAAPTTPKEPATTTKBPATTTKBP 840
QY 798 KPAPTTPETPTTSEVSTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEGVPT 857
Db 841 KPAPTTPETPTTSEVSTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEGVPT 900
QY 858 TKTPAATKEPMITAKDKTTERDLRTTPETTTAAPKMTKETAATTEKTTESKITATTTQV 917
Db 901 TKTPAATKEPMITAKDKTTERDLRTTPETTTAAPKMTKETAATTEKTTESKITATTTQV 960
QY 918 TSTTTQDTPFKITTLTKTTLAPKVTITTKTITTEIMNKPEETAKPKBATNSKATTPK 977
Db 961 TSTTTQDTPFKITTLTKTTLAPKVTITTKTITTEIMNKPEETAKPKBATNSKATTPK 1020
QY 978 PQKPTKAPKPTSTKPKTMVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRPN 1037
Db 1021 PQKPTKAPKPTSTKPKTMVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRPN 1080
QY 1038 QTPNSKLVEVPKSDAGGAGETPHMLLRPHVFMPEVTPDMYLLPRVNOGIIINPMLS 1097
Db 1081 QTPNSKLVEVPKSDAGGAGETPHMLLRPHVFMPEVTPDMYLLPRVNOGIIINPMLS 1140
QY 1098 DETNICNGKVPDGLTTLRNGTLVAFRGHVFWMLSPSPSPAPRITTEVWGIPSPIDVT 1157
Db 1141 DETNICNGKVPDGLTTLRNGTLVAFRGHVFWMLSPSPSPAPRITTEVWGIPSPIDVT 1200
QY 1158 RCNCEGKTFKFDOSQWRRFTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPEVY 1217
Db 1201 RCNCEGKTFKFDOSQWRRFTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPEVY 1260
QY 1218 FFKRGGSIOQYIYKQBPVKCGRRPALNYPVYVGMTOVRRRFRFAIGPSQTHIRIQY 1277
Db 1261 FFKRGGSIOQYIYKQBPVKCGRRPALNYPVYVGMTOVRRRFRFAIGPSQTHIRIQY 1320
QY 1278 SPARLAYQDKGVLHNEVKVYSILWRGIPNVVVTSAISLPNIRKPDGDYDYAFSKDQYNNIDV 1337
Db 1321 SPARLAYQDKGVLHNEVKVYSILWRGIPNVVVTSAISLPNIRKPDGDYDYAFSKDQYNNIDV 1380

QY 1338 PSRTARAITTRSGQTLSKWYNCP 1361
Db 1381 PSRTARAITTRSGQTLSKWYNCP 1404
RESULT 5
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Iuann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 99.8%; Score 7308.7; DB 13; Length 1404;
Best Local Similarity 96.9%; Pred No. 3.9e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KEVCTAELSCRCRCESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
Db 61 KEVCTAELSCRCRCESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQITKSTTKRSPKPPNKKTKVIESPEITE----- 156
Db 121 PPSGASQITKSTTKRSPKPPNKKTKVIESPEITE----- 156

QY	157	-----VKONKQRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST	197
Db	181	KIKSSKNSAANRELQKLLKVKONKQRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST	240
QY	198	TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKETTVETKETTINKQSTDG	257
Db	241	TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKETTVETKETTINKQSTDG	300
QY	258	KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALITPKETPTTPKPEPAS	317
Db	301	KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALITPKETPTTPKPEPAS	360
QY	318	TTPKETPTTTIKSAPTTKBPAPTTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	377
Db	361	TTPKETPTTTIKSAPTTKBPAPTTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	420
QY	378	APTTSKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	437
Db	421	APTTSKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	480
QY	438	EPAPTAPKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	497
Db	481	EPAPTAPKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	540
QY	498	TTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	557
Db	541	TTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	600
QY	558	APTAPKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	617
Db	601	APTAPKBPAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP	660
QY	618	PEEPAPTTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	677
Db	661	PEEPAPTTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	720
QY	678	APTTPKBPAPKELAPTTKBPAPTTSDKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	737
Db	721	APTTPKBPAPKELAPTTKBPAPTTSDKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	780
QY	738	TAPTTLKBPAPTTKBPAPKELAPTTTKGPTSTSDKBPAPTTKBPAPTTKBPAPTTKBP	797
Db	781	TAPTTLKBPAPTTKBPAPKELAPTTTKGPTSTSDKBPAPTTKBPAPTTKBPAPTTKBP	840
QY	798	KPAPTTPETPTPTSEVSTPTTKPTTIHKSPDESTPELSAETTPKALENSPKPEGVPT	857
Db	841	KPAPTTPETPTPTSEVSTPTTKPTTIHKSPDESTPELSAETTPKALENSPKPEGVPT	900
QY	858	TKTPAAATKPEMTTAKDKTTERDARTPETTTAAPKMTKETATTTKTESKITATTTQV	917
Db	901	TKTPAAATKPEMTTAKDKTTERDARTPETTTAAPKMTKETATTTKTESKITATTTQV	960
QY	918	TSITTTQDPTPKITTLTKTTTLAPKVTTTKKTIITTEIMNKPEETAAPKDRATNSKATTPK	977
Db	961	TSITTTQDPTPKITTLTKTTTLAPKVTTTKKTIITTEIMNKPEETAAPKDRATNSKATTPK	1020
QY	978	POKPTKAPKPTSTKPKTPMVRKPTTTPRKMSTWPEINPTSRIAEAMLTQTTTPN	1037
Db	1021	POKPTKAPKPTSTKPKTPMVRKPTTTPRKMSTWPEINPTSRIAEAMLTQTTTPN	1080
QY	1038	QTPNSKLVBNPKSEDAGAGETPHMLLRHVEMPEVTPDMXYLPRVNOGIIINPMLS	1097
Db	1081	QTPNSKLVBNPKSEDAGAGETPHMLLRHVEMPEVTPDMXYLPRVNOGIIINPMLS	1140
QY	1098	DETNIKNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSIDTFT	1157
Db	1141	DETNIKNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSIDTFT	1200
QY	1158	RCNCEGKTTFFKDSQVWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPEVY	1217
Db	1201	RCNCEGKTTFFKDSQVWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPEVY	1260

RESULT 6
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match 96.7%; Score 7084.3; DB 13; Length 1354;
Best Local Similarity 97.0%; Pred. No. 1e-163;
Matches 1320; Conservative 8; Mismatches 26; Indels 7; Gaps 1;
QY 1 MAWKTLPIYLLILLISVFIQVSSQDLSSCAGRCGEGYSRDTNCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLILLISVFIQVSSQDLSSCAGRCGEGYSRDTNCNDYNCQHYMECCPDF 60

RESULT 7

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
UTS-10-134-557-AC

```

Query Match 96.4%; Score 7058.9; DB 13; Length 1320;
Best Local Similarity 97.0%; Pred. No. 4.1e-163;
Matches 1320; Conservative 0; Mismatches 0; Indels 41;

QY 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGRCGSGYSRDATCNCYNCQHMECCPDF 60
Db 1 MAWKTLPIVILLLLSVFVIQQVSSQ-----25
QY 61 KRVCFAELSCGRCFESFERGRCDCAQCKKYDKCCPDYESFAEVHNPTSPSSKKAP 120
Db 26 -----ELSCGRCFESFERGRCDCAQCKKYDKCCPDYESFAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEVKONKXRTKKKPTPKPPVDEAG 180
Db 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEVKONKXRTKKKPTPKPPVDEAG 139
QY 181 SGLDNGDFKVTTPDTSTQHNVKYSTSPKITTAKPINRPSLPNSOTSKESTJTVNKETT 240
Db 140 SGLDNGDFKVTTPDTSTQHNVKYSTSPKITTAKPINRPSLPNSOTSKESTJTVNKETT 199
QY 241 VETKETTTNNKQSTDGKEKTTSAKETQSIETSAKDLAPTSKVLAKPTPKAETTTKGPA 300
Db 200 VETKETTTNNKQSTDGKEKTTSAKETQSIETSAKDLAPTSKVLAKPTPKAETTTKGPA 259
QY 301 LTTPKBPTTTPKEPASTTPKEBPTTTIKGAPTTPKBPAPTTPPKSAPTTPKBPAPTTTKE 360
Db 260 LTTPKBPTTTPKEPASTTPKEBPTTTIKGAPTTPKBPAPTTPPKSAPTTPKBPAPTTTKE 319
QY 361 PAPTTPKEPAPTTTKEPAPTTTKSAPTTPKBPAPTTPPKBPAPTTPKEBPTT 420
Db 320 PAPTTPKEPAPTTTKEPAPTTTKSAPTTPKBPAPTTPPKBPAPTTPKEBPTT 379
QY 421 PKBPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKPSPTTKEPA 480
Db 380 PKBPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKPSPTTKEPA 439
QY 481 PTTTKSAPTTTKEPAPTTTKSAPTTPKBPAPTTPKEPAPTTPKEPAPTTPKBPAPTTPKE 540
Db 440 PTTTKSAPTTTKEPAPTTTKSAPTTPKBPAPTTPKEPAPTTPKEPAPTTPKBPAPTTPKE 499
QY 541 PAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKAPTTPKLTPTTPEKLAPTTPKEPAPT 600
Db 500 PAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKAPTTPKLTPTTPEKLAPTTPKEPAPT 559
QY 601 TPELAPTTPKEPTTTPKEPAPTTPKAAAPNTPKBPAPTTPKEPAPTTPKEPAPTTPKE 660
Db 560 TPELAPTTPKEPTTTPKEPAPTTPKAAAPNTPKBPAPTTPKEPAPTTPKEPAPTTPKE 619
QY 661 TAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTPKGTSTTSKBPAPTTPKGTAPTTPK 720
Db 620 TAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTPKGTSTTSKBPAPTTPKGTAPTTPK 679
QY 721 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTPKGTSTTSKBPAPTTP 780
Db 680 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTPKGTSTTSKBPAPTTP 739
QY 781 KETAPTTPKEPAPTTPKBPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 840
Db 740 KETAPTTPKEPAPTTPKBPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 799
QY 841 PTPKALENSPKBPVPTTKTTPAAKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 900
Db 800 PTPKALENSPKBPVPTTKTTPAAKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETAT 859
QY 901 TTEKTTESKITATTTQVNTSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 960
Db 860 TTEKTTESKITATTTQVNTSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 919
QY 961 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMRVRKPTTTPPKMTSTMPBLN 1020
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMRVRKPTTTPPKMTSTMPBLN 979
QY 1021 PTSRIAEMLQTTTPNQTPNSKLVVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 980 PTSRIAEMLQTTTPNQTPNSKLVVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1039

1081 YLPRVNOGIIINPMLSDETNICNGKPVLDGLTLTRNGTLVAFRGHYFWMLSPFSPSPAR 1140
1040 YLPRVNOGIIINPMLSDETNICNGKPVLDGLTLTRNGTLVAFRGHYFWMLSPFSPSPAR 1099
1141 RTEVWGIFSPIDTVTRCNCCEGKTFPFKDSQVWRFTNDIKDAGYKPKIPKGFGLTGOI 1200
1100 RTEVWGIFSPIDTVTRCNCCEGKTFPFKDSQVWRFTNDIKDAGYKPKIPKGFGLTGOI 1159
1201 VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBPVOKPCGRRPALNVVYVGMTQVRRRR 1260
1160 VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBPVOKPCGRRPALNVVYVGMTQVRRRR 1219
1261 FERAIGPSQHTTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1320
1220 FERAIGPSQHTTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1279
1321 GYDYAFSKDQYYNIDVPSRTARAITRSGQTLISKVWYNCP 1361
1280 GYDYAFSKDQYYNIDVPSRTARAITRSGQTLISKVWYNCP 1320

RESULT 8
US-10-124-557-60
; Sequence 60, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 96.4%; Score 7058.9; DB 13; Length 1320;
 Best Local Similarity 97.0%; Pred. No. 4, 1e-163;
 Matches 1320; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSODLSSCAGGEGYSDATCNCYHMECCPDF 60
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

QY 61 KRVCATLSCGRCPESPERGECDCQCKYKCCPDYSEFCAEVNPTSPSSKKAP 120
 DB -----ELSCGRCPESPERGECDCQCKYKCCPDYSEFCAEVNPTSPSSKKAP 79

QY 121 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITEVDKNKRNKTKKPTPKPPVWDEAG 180
 DB 80 PPASGASQIKSTTKRSPKPNKKTKKVIIESEITEVDKNKRNKTKKPTPKPPVWDEAG 139

QY 181 SGLDNGDFKVTPTDSTTQHNKVTSPKITTAKPINPRPSLPNDSKETSITVKNKETT 240
 DB 140 SGLDNGDFKVTPTDSTTQHNKVTSPKITTAKPINPRPSLPNDSKETSITVKNKETT 199

QY 241 VETKEITTNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 300
 DB 200 VETKEITTNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 259

QY 301 LTTPEKPTPTPKPEASTTPEKPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360
 DB LTTPEKPTPTPKPEASTTPEKPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 319

QY 361 PAPTTPKEAPTTTKEAPATTTKSAPTTPKPEAPTTPKKAPATTTPKPEAPTTTKEPTPT 420
 DB 320 PAPTTPKEAPTTTKEAPATTTKSAPTTPKPEAPTTPKKAPATTTPKPEAPTTTKEPTPT 379

QY 421 PKPEAPTTKPEAPTTKPEAPTAPEKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKEPA 480
 DB 380 PKPEAPTTKPEAPTTKPEAPTAPEKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKEPA 439

QY 481 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEAPTTTKEAPTTKPEAPTTKPEAPTTTKE 540
 DB 440 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEAPTTTKEAPTTKPEAPTTTKEAPTTTKE 499

QY 541 PAPTTPKEAPTTTKKAPTAPEKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKEKAPT 600
 DB 500 PAPTTPKEAPTTTKKAPTAPEKAPTTKPEAPTTKPEAPTTKPEAPTTTKEKAPT 559

QY 601 TPBELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTTKE 660
 DB 560 TPBELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTTKE 619

QY 661 TAPTTPKGTAPTTLKEAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTTPK 720
 DB 620 TAPTTPKGTAPTTLKEAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTTPK 679

QY 721 EPAPTTPKPEAPTTKGTAPTTLKEAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTP 780
 DB 680 EPAPTTPKPEAPTTKGTAPTTLKEAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTP 739

QY 781 KETAPTTKPEAPTTPKKAPATTPEPTPTTSEVSTPTTKEPTTIHKSPDESTPBLSAE 840
 DB 740 KETAPTTKPEAPTTPKKAPATTPEPTPTTSEVSTPTTKEPTTIHKSPDESTPBLSAE 799

QY 841 PTPKALENPKPEGVPTTKTPAATKPEMTTAKDKTTTERDLRTPPTTTAAAPKMTKETAT 900
 DB 800 PTPKALENPKPEGVPTTKTPAATKPEMTTAKDKTTTERDLRTPPTTTAAAPKMTKETAT 859

QY 901 TTEKTTESKITTATTQVSTTTQDTPPKITLTKTTTLAPKVTTTKKITTITTEIMNKPBE 960
 DB 860 TTEKTTESKITTATTQVSTTTQDTPPKITLTKTTTLAPKVTTTKKITTITTEIMNKPBE 919

QY 961 TAKPKDRATNSKATTPKPKOKPTKAPKPKPTSTKKPKTMPVRPKPTTTPPKMTSTNPELN 1020
 DB 920 TAKPKDRATNSKATTPKPKOKPTKAPKPKPTSTKKPKTMPVRPKPTTTPPKMTSTNPELN 979

QY 1021 PTSRTAEAMLQTTTRPNQTPNSKLVVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1080
 DB 980 PTSRTAEAMLQTTTRPNQTPNSKLVVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1039

QY 1081 YLPRVFNQGIILNPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1140
 DB 1040 YLPRVFNQGIILNPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1099

QY 1141 RITEVWGPSPIDTVFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKIFKGFGLTGOI 1200
 DB 1100 RITEVWGPSPIDTVFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKIFKGFGLTGOI 1159

QY 1201 VAALSTAKYKNWPSVYFFKKGSSIQVIYKQBPVKCPCGRRPALNYPVYGEMTQVRRRR 1260
 DB 1160 VAALSTAKYKNWPSVYFFKKGSSIQVIYKQBPVKCPCGRRPALNYPVYGEMTQVRRRR 1219

QY 1261 FERAIGPSQTHTRI QXSPARLAYQDKGLVHNEVKVSLILWRGLNVVTSALSLNIRKPD 1320
 DB 1220 FERAIGPSQTHTRI QXSPARLAYQDKGLVHNEVKVSLILWRGLNVVTSALSLNIRKPD 1279

QY 1321 GYDYAFSKQYINIDVPSTRARAITTRSGOTLSKVWYNCP 1361
 DB 1280 GYDYAFSKQYINIDVPSTRARAITTRSGOTLSKVWYNCP 1320

RESULT 9

US-10-124-557-42
 ; Sequence 42, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 42:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1311 amino acids
 ; TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match      96.3%; Score 7049; DB 13; Length 1311;
Best Local Similarity 96.3%; Pred. No. 7e-163; 0; Indels 50; Gaps 1;
Matches 1311; Conservative 0; Mismatches 0;

QY 1 1MAWKTLPIYLLLSVFEVIOQVSSQDLSSCAGRCGEGYSRDATCNCNDYNOQHMECCPDF 60
Db 1 1MAWKTLPIYLLLSVFEVIOQVSSQDLSSCAGRCGEGYSRDATCNCNDYNOQHMECCPDF 60
QY 61 KRVTAEKSCGRCFESFERGRCDCDAQCKYDKCCPDYSECAEVHNFTSPSSKAP 120
Db 61 KRVTAEKSCGRCFESFERGRCDCDAQCKYDKCCPDYSECAEVHNFTSPSSKAP 120
QY 61 KRVTAEKSCGRCFESFERGRCDCDAQCKYDKCCPDYSECAEVHNFTSPSSKAP 120
Db 61 KRVTAEKSCGRCFESFERGRCDCDAQCKYDKCCPDYSECAEVHNFTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEVKDNKKNTKKKPTKPPVVDKAP 180
Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEVKDNKKNTKKKPTKPPVVDKAP 180
QY 106 -----EVDKNNKNTKKKPTKPPVVDKAP 130
Db 106 -----EVDKNNKNTKKKPTKPPVVDKAP 130
QY 181 SGLDNGDFKVTPTDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 240
Db 181 SGLDNGDFKVTPTDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 240
QY 131 SGLDNGDFKVTPTDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 190
Db 131 SGLDNGDFKVTPTDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 190
QY 241 VETKETTNTKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 300
Db 241 VETKETTNTKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 300
QY 191 VETKETTNTKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 250
Db 191 VETKETTNTKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 250
QY 301 LTTPKETPTTKPEASPTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPE 360
Db 301 LTTPKETPTTKPEASPTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPE 360
QY 251 LTTPKETPTTKPEASPTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPE 310
Db 251 LTTPKETPTTKPEASPTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPE 310
QY 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPT 420
Db 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPT 420
QY 311 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPT 370
Db 311 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTPT 370
QY 421 PKPEAPTTKEAPTTTKPEAPTTAPKAPAPTTPKPEAPTTKEAPTTTKPEAPTTKEPE 480
Db 421 PKPEAPTTKEAPTTTKPEAPTTAPKAPAPTTPKPEAPTTKEAPTTTKPEAPTTKEPE 480
QY 371 PKPEAPTTKEAPTTTKPEAPTTAPKAPAPTTPKPEAPTTKEAPTTTKPEAPTTKEPE 430
Db 371 PKPEAPTTKEAPTTTKPEAPTTAPKAPAPTTPKPEAPTTKEAPTTTKPEAPTTKEPE 430
QY 481 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEPTTKSAPTTTKPEAPTTTKPKAPPTPKPE 540
Db 481 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEPTTKSAPTTTKPEAPTTTKPKAPPTPKPE 540
QY 431 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEPTTKSAPTTTKPEAPTTTKPKAPPTPKPE 490
Db 431 PTTTKSAPTTTKPEAPTTTKSAPTTPKPEPTTKSAPTTTKPEAPTTTKPKAPPTPKPE 490
QY 541 PAPTTPKEAPTTTKKAPAPTTAPKAPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPT 600
Db 541 PAPTTPKEAPTTTKKAPAPTTAPKAPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPT 600
QY 491 PAPTTPKEAPTTTKKAPAPTTAPKAPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPT 550
Db 491 PAPTTPKEAPTTTKKAPAPTTAPKAPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPT 550
QY 601 TPEELAPTTPEPTPTTPEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 660
Db 601 TPEELAPTTPEPTPTTPEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 660
QY 551 TPEELAPTTPEPTPTTPEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 610
Db 551 TPEELAPTTPEPTPTTPEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 610
QY 661 TAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKGTAPTTPK 720
Db 661 TAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKGTAPTTPK 720
QY 611 TAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKGTAPTTPK 670
Db 611 TAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKGTAPTTPK 670
QY 721 EPAPTTPKEAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPK 780
Db 721 EPAPTTPKEAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPK 780
QY 671 EPAPTTPKEAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPK 730
Db 671 EPAPTTPKEAPTTPKGTATTLKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPK 730
QY 781 KETAPTTPKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKKAPKELAPTTTKE 840
Db 781 KETAPTTPKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKKAPKELAPTTTKE 840
QY 731 KETAPTTPKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKKAPKELAPTTTKE 790
Db 731 KETAPTTPKEAPTTPKKAPKELAPTTTKEPTSTTSOKPAPTTPKKAPKELAPTTTKE 790
QY 841 PTPKALENSPKGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
Db 841 PTPKALENSPKGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
QY 791 PTPKALENSPKGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 850
Db 791 PTPKALENSPKGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 850
QY 901 TTEKTTESKITAATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTTKTIITTEIMNKPEE 960
Db 901 TTEKTTESKITAATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTTKTIITTEIMNKPEE 960
QY 851 TTEKTTESKITAATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTTKTIITTEIMNKPEE 910
Db 851 TTEKTTESKITAATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTTKTIITTEIMNKPEE 910

RESULT 10
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
```


FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match 93.4%; Score 6836.3; DB 13; Length 1314;
Best Local Similarity 94.1%; Pred. No. 1e-157;
Matches 1281; Conservative 8; Mismatches 25; Indels 47; Gaps 2;

QY 1 MAWKTLPIYLLILLVSVFVQQVSSQDLSSCAGRCGEGYSRDATCNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLILLVSVFVQQVSSQDLSSCAGRCGEGYSRDATCNCQHYMECCPDF 60

QY 61 KRVCYAEISCKGRCFESFERGECDDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVCYAEISCKGRCFESFERGECDDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

QY 121 PPSGASQTIKSTKRSKPPNKKTKVIBSEBITEVKONKNRTKKKPTKPPVVDKAP 180
DB 81 SSSSSSTTWIKSSXNSAANRELQKL-----KVONKNRTKKKPTKPPVVDKAP 133

QY 181 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 240
DB 134 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 193

QY 241 VETKETTNTKQSTDKKETTSAKETQSTKTSADKLAPTSKVLAKPTPKAETTTKGA 300
DB 194 VETKETTNTKQSTDKKETTSAKETQSTKTSADKLAPTSKVLAKPTPKAETTTKGA 253

QY 301 LTTTPKEPTTPKPEASTTKEPTTTIKSAPTTKPEAPTTTKKSAPTTKPEAPTTTKE 360
DB 254 LTTTPKEPTTPKPEASTTKEPTTTIKSAPTTKPEAPTTTKKSAPTTKPEAPTTTKE 313

QY 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKPAPTTKPEAPTTKPEPTPT 420
DB 314 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKPAPTTKPEAPTTKPEPTPT 373

QY 421 PKPEAPTTKPEAPTTKPEAPTTAPKPAPTTKPEAPTTKPEAPTTTKKPSPTTKPEA 480
DB 374 PKPEAPTTKPEAPTTKPEAPTTAPKPAPTTKPEAPTTKPEAPTTTKKPSPTTKPEA 433

QY 481 PTTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTTKPE 540
DB 434 PTTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTTKPE 493

QY 541 PAPTTPKEAPTTTKKPAPTAPKPAPTTKPEAPTTPKKLTPPTPKLAPTTPEKAPPT 600
DB 494 PAPTTPKEAPTTTKKPAPTAPKPAPTTKPEAPTTPKKLTPPTPKLAPTTPEKAPPT 553

QY 601 TPELAPTTPEPTPTTPEEAPTTPKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKE 660
DB 554 TPELAPTTPEPTPTTPEEAPTTPKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKE 613

QY 661 TAPTTPKGTATTLKPEAPTTKPKAPKELAPTTTKETSTTSKAPPTTKGAPTTTPK 720
DB 614 TAPTTPKGTATTLKPEAPTTKPKAPKELAPTTTKETSTTSKAPPTTKGAPTTTPK 673

QY 721 EPAPTTPKEAPTTTKGAPTTTLKPEAPTTTKKPAKPELAPTTTKGPTSTTSKAPPTTP 780
DB 674 EPAPTTPKEAPTTTKGAPTTTLKPEAPTTTKKPAKPELAPTTTKGPTSTTSKAPPTTP 733

QY 781 KETAPTTPKEAPTTTPKPAPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 840
DB 734 KETAPTTPKEAPTTTPKPAPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 793

QY 841 PTPKALENSPKPEPGVPTTKTAAATKPEMTTAAKDKTTERDLRTTETTTAAAPKMTKETAT 900
DB 794 PTPKALENSPKPEPGVPTTKTAAATKPEMTTAAKDKTTERDLRTTETTTAAAPKMTKETAT 853

QY 901 TTEKTESKITAATTQVSTTTQDTPPKITTLTKTTLAPKVTTKTKTTTTEIMNKPEE 960
DB 854 TTEKTESKITAATTQVSTTTQDTPPKITTLTKTTLAPKVTTKTKTTTTEIMNKPEE 913

QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMTPEVRKPKTTPTRKMTSTMPELN 1020
DB 914 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMTPEVRKPKTTPTRKMTSTMPELN 973

QY 1021 PITSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPVEVTPDMD 1080
DB 974 PITSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPVEVTPDMD 1033

QY 1081 YLPRVNPQGIILNPMLSDETNICNGKPVLDGLTLRLNGTLVAPRGHYFWMLSPEPSPAR 1140
DB 1034 YLPRVNPQGIILNPMLSDETNICNGKPVLDGLTLRLNGTLVAPRGHYFWMLSPEPSPAR 1093

QY 1141 RITEVWGIPIPDITVTRCNCEGKTPFFKDSQVWRFNDIKDAGYKPIFKGFGGUTGQI 1200
DB 1094 RITEVWGIPIPDITVTRCNCEGKTPFFKDSQVWRFNDIKDAGYKPIFKGFGGUTGQI 1153

QY 1201 VAALSTAKYNKPNPESVYFFKRGSGIOQYIKQEPVQKCPGRPALNYPVYVGMTQVRRR 1260
DB 1154 VAALSTAKYNKPNPESVYFFKRGSGIOQYIKQEPVQKCPGRPALNYPVYVGMTQVRRR 1213

QY 1261 FERAIQSPQTHIRIQYSPARLAYQDKGLHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1320
DB 1214 FERAIQSPQTHIRIQYSPARLAYQDKGLHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1273

QY 1321 GYDYAFSKDQYNNIDVPSRTARAITTRSGOTLSKVWYNCP 1361
DB 1274 GYDYAFSKDQYNNIDVPSRTARAITTRSGOTLSKVWYNCP 1314

RESULT 12
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserit, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 142:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1313 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
 US-10-124-557-142

Query Match 93.1%; Score 6820.2; DB 13; Length 1313;
 Best Local Similarity 94.0%; Pred. No. 2.5e-157;
 Matches 1279; Conservative 8; Mismatches 26; Indels 48; Gaps 2;

```

QY 1 MAWKTLPIYLLLSVFIQOVSSQDLSSCAGRCGEGYSDATCNCYNCQHVMCCPDF 60
DB 1 MAWKTLPIYLLLSVFIQOVSSQ-----25
QY 61 KRVCETLSCKGCFEFREGRECDCAQCKYDKCCPDYSEFCAEVNHTSPSSKAP 120
DB 26 -----ELSCGKCFEFREGRECDCAQCKYDKCCPDYSEFCAEHSVSENQSSSS 79
QY 121 PPSGASQIKSTTKRSPKPKKTKKXKXIESEITEVKDNKTKKTKKPKPPVDEAG 180
DB 80 SSSSSSSTLIWKSSXNSAANRELQKL-----KVDKNKTKKTKKPKPPVDEAG 132
QY 181 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETT 240
DB 133 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETT 192
QY 241 VETKEITTNKQSTDGKKEITSAKETOSIEKTSKADLAPTSKVLAKEPTKAPTITKGA 300
DB 193 VETKEITTNKQSTDGKKEITSAKETOSIEKTSKADLAPTSKVLAKEPTKAPTITKGA 252
QY 301 LTPKKEPTTTPKEPASTTPKEPTTIKSAPTTPKEPATTTPKSAPTTPKEPATTTPKE 360
DB 253 LTPKKEPTTTPKEPASTTPKEPTTIKSAPTTPKEPATTTPKSAPTTPKEPATTTPKE 312
QY 361 PAPTTPKEPATTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKEPTPT 420
DB 313 PAPTTPKEPATTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKEPTPT 372
QY 421 PKEPATTTPKEPATTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKEPA 480
DB 373 PKEPATTTPKEPATTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKEPA 432
QY 481 PTTTTSAPTTTKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 540
DB 433 PTTTTSAPTTTKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 492
QY 541 PAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPAPT 600
DB 493 PAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPAPT 552
QY 601 TPEELAPTTPEEPTTTPKEPATTTPKAAAPNTPEPATTTPKSAPTTPKEPATTTPKE 660
DB 553 TPEELAPTTPEEPTTTPKEPATTTPKAAAPNTPEPATTTPKSAPTTPKEPATTTPKE 612
QY 661 TAPTTPKGTAPTTLKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPK 720
DB 613 TAPTTPKGTAPTTLKEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPK 672
  
```

```

QY 721 EPAPTTPEPATTTPKGTAPTTLKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 780
DB 673 EPAPTTPEPATTTPKGTAPTTLKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 732
QY 781 KETAPTTPEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 840
DB 733 KETAPTTPEPATTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 792
QY 841 PTPKALENSPKPEGVPTTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 900
DB 793 PTPKALENSPKPEGVPTTTPKSAPTTPKEPATTTPKSAPTTPKEPATTTPKSAPTTPKE 852
QY 901 TTEKTTESKITATTTOVTSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 960
DB 853 TTEKTTESKITATTTOVTSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 912
QY 961 TAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 1020
DB 913 TAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 972
QY 1021 PTSRIABAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD 1080
DB 973 PTSRIABAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD 1032
QY 1081 YLPRVNPQGIILNPMLSDETNI CNKPKVDGLTTLRNGLTVAFRGHYFWMLSPPSPSPAR 1140
DB 1033 YLPRVNPQGIILNPMLSDETNI CNKPKVDGLTTLRNGLTVAFRGHYFWMLSPPSPSPAR 1092
QY 1141 RITEVWGIPIPIDTFTRCNCEGKTFPKKSGSIQYIYKQBPVQKCGRRPALNYPVVGEMTQVRRR 1260
DB 1093 RITEVWGIPIPIDTFTRCNCEGKTFPKKSGSIQYIYKQBPVQKCGRRPALNYPVVGEMTQVRRR 1212
QY 1201 VAALSTAKYKNWPSVYFVKRGGSIQYIYKQBPVQKCGRRPALNYPVVGEMTQVRRR 1320
DB 1153 VAALSTAKYKNWPSVYFVKRGGSIQYIYKQBPVQKCGRRPALNYPVVGEMTQVRRR 1272
QY 1261 FERAIGPSQHTTIRIQYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1320
DB 1213 FERAIGPSQHTTIRIQYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1272
QY 1321 GYDYAFSKDQYINIDVPSTARAITTRSGOTLSKVWYNCP 1361
DB 1273 GYDYAFSKDQYINIDVPSTARAITTRSGOTLSKVWYNCP 1313
  
```

RESULT 13

US-10-124-557-44
 ; Sequence 44, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Geener, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/643,502
;; FILING DATE: 18-JAN-1991
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cserr, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 876-1170
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 44:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1270 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match 92.7%; Score 6784.9; DB 13; Length 1270;
Best Local Similarity 93.3%; Pred. No. 1.7e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

Qy	1	MAWKTLPIYLLLSVFIQVVSODLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF	60
Db	1	MAWKTLPIYLLLSVFIQVVSQ-----	25
Qy	61	KRVCTAELCKGRCFESFERGECDDAQCKYDKCCPDYESFCAEVHNETSPSSKAP	120
Db	26	-----ELSKGRCFESFERGECDDAQCKYDKCCPDYESFCA-----	64
Qy	121	PPGASQTIKSTTKRSPKPNKXKTKVIESEETEVKDNKNKRTKKKPKPPVDEAG	180
Db	65	-----EVKDNKNKRTKKKPKPPVDEAG	89
Qy	181	SLDNGDFKVTPTDSTQHNKVTSPKITTAKPINRPSLPNSDTSKETSILVNKETT	240
Db	90	SLDNGDFKVTPTDSTQHNKVTSPKITTAKPINRPSLPNSDTSKETSILVNKETT	149
Qy	241	VETKETTITNKQTSIDGKEKITSASKEISIEKTSAKOLAPTSKVLAKPTPKAETTKGPA	300
Db	150	VETKETTITNKQTSIDGKEKITSASKEISIEKTSAKOLAPTSKVLAKPTPKAETTKGPA	209
Qy	301	LTTPKEPTPTTKGPASTTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE	360
Db	210	LTTPKEPTPTTKGPASTTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE	269
Qy	361	PAPTPKPEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKPAPTTKPEAPTTKPEPTTT	420
Db	270	PAPTPKPEAPTTTKPEAPTTTKSAPTTKPEAPTTPKKPAPTTKPEAPTTKPEPTTT	329
Qy	421	PKEAPTTKPEAPTTTKPEAPTTAPKKPAPTTKPEAPTTKPEAPTTTKKPSPTTKPEPA	480
Db	330	PKEAPTTKPEAPTTTKPEAPTTAPKKPAPTTKPEAPTTKPEAPTTTKKPSPTTKPEPA	389
Qy	481	PTTKSAPTTKPEAPTTTKSAPTTTKPEPTTKPEAPTTKPEAPTTKPEAPTTKPEPTKE	540
Db	390	PTTKSAPTTKPEAPTTTKSAPTTTKPEPTTKPEAPTTKPEAPTTKPEAPTTKPEPTKE	449
Qy	541	PAPTPKPEAPTTTKKPAATAKPEAPTTPKETAPTPKKLTPTTPEKLAPTTPKPAPT	600
Db	450	PAPTPKPEAPTTTKKPAATAKPEAPTTPKETAPTPKKLTPTTPEKLAPTTPKPAPT	509
Qy	601	TPEELAPTTPEPTPTTPEAPTTTPKAAANPTKPEAPTTKPEAPTTKPEAPTTTKE	660
Db	510	TPEELAPTTPEPTPTTPEAPTTTPKAAANPTKPEAPTTKPEAPTTKPEAPTTTKE	569

Qy	661	TAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPK	720
Db	570	TAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPK	629
Qy	721	EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTP	780
Db	630	EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTP	689
Qy	781	KETAPTTPKEPAPTTPKAPPTTPETPTTSEVSTPTTKEPTTHKSPDESTPELSAE	840
Db	690	KETAPTTPKEPAPTTPKAPPTTPETPTTSEVSTPTTKEPTTHKSPDESTPELSAE	749
Qy	841	PTPKALENSPKBPGVPTTKTTPAATKPEMTTAKOKTTERDLRTPPTTAAPOMTKETAT	900
Db	750	PTPKALENSPKBPGVPTTKTTPAATKPEMTTAKOKTTERDLRTPPTTAAPOMTKETAT	809
Qy	901	TTEKTTESKITATTQVTSITQDTTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEE	960
Db	810	TTEKTTESKITATTQVTSITQDTTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEE	869
Qy	961	TAKPKDRATNSKATTPKPOKTPKAPKPTSTKKPKTMPRVKPKTTPPKMTSTMPELN	1020
Db	870	TAKPKDRATNSKATTPKPOKTPKAPKPTSTKKPKTMPRVKPKTTPPKMTSTMPELN	929
Qy	1021	PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD	1080
Db	930	PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD	989
Qy	1081	YLPKRVNPQGIINPMLSDETNICNGRPVDGLTTLRNGTLVAFRGHVFWMLSPPSPSPAR	1140
Db	990	YLPKRVNPQGIINPMLSDETNICNGRPVDGLTTLRNGTLVAFRGHVFWMLSPPSPSPAR	1049
Qy	1141	RITVEVGIPIPIDTFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI	1200
Db	1050	RITVEVGIPIPIDTFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI	1109
Qy	1201	VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRR	1260
Db	1110	VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRR	1169
Qy	1261	FERAIGSQTHTRIQYSPARLAYQDKGVLHNEKVSILWGLEPNVVTSAISLNPNRKPD	1320
Db	1170	FERAIGSQTHTRIQYSPARLAYQDKGVLHNEKVSILWGLEPNVVTSAISLNPNRKPD	1229
Qy	1321	GIDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKVMYNCP	1361
Db	1230	GIDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKVMYNCP	1270

RESULT 14
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 74.9%; Score 5485.7; DB 13; Length 1038;
Best local Similarity 76.3%; Pred. No. 4.2e-125; Indels 323; Gaps 3;
Matches 1038; Conservative 0; Mismatches 0

1 MAWKTLPIYLLLLSVFVIQQVSSQLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
61 KRVCUTAEISCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
26 -----ELSCCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCA----- 64
121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEEITEVKNKNTKKKPTPKPPVDEAG 180
65 -----EVKNKNTKKKPTPKPPVDEAG 89
181 SGLDNGDPKVTTPDTSITQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKETT 240
90 SGLDNGDKVTPDTSITQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKETT 149
241 VETKETTNNKQSTGDKGKETSASAKTOSIEKTSKOLAPTSKULAKETPKAEITTKGPA 300
150 VETKETTNNKQSTGDKGKETSASAKTOSIEKTSKOLAPTSKULAKETPKAEITTKGPA 209
301 LTTKEPNTTPKEPASTTPKEPTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKE 360
210 LTTKEPNTTPKEPASTTPKEPTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKE 269
361 PAPTTTKEPAPTTTKEPAPTTTKSAPTTKKEPAPTTPKKAPTTTKEPAPTTTKEPTPTT 420
270 PAPTTTKEPAPTTTKEPAPTTTKSAPTTKKEPAPTTPKKAPTTTKEPAPTTTKEPTPTT 329
421 PKEPAPTTTKEPAPTTTKEPAPTTAPKKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480
330 PKEPAPTTTKEPAPTTTKEPAPTTAPKKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 389
481 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
390 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 449
541 PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600
450 PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 509
601 TPEELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660
510 TPEELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 569
661 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720
570 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 629
721 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 780
630 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 689

Search completed: October 13, 2004, 11:53:23
Job time : 146.466 secs

SEQUENCE FROM N.A.
Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
Kelleher K., Glanvott J., Calvetti J., Fitzgerald M., Kriz M.J.,
Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
Bhatia S., Kriz R., Hewick R., Clark S.C.;
Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
EMBL; U70136; AAB09089.1; -.
HSP; P04004; LOC0.
Genew; HGNC:9364; PRG4.
GO; GO:0008283; P:cell proliferation; TAS.
InterPro; IPR000585; Hemopexin.
InterPro; IPR001212; Somatomedin_B.
Pfam; PF00045; Hemopexin; 2.
Pfam; PF01033; Somatomedin_B; 2.
PRINTS; PR00022; SOMATOMEDINB.
SMART; SM00120; HX; 2.
SMART; SM00201; SO; 2.
PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
PROSITE; PS00524; SOMATOMEDIN_B; 2.
SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64.

Query Match		99.8%;	Score 7308.7;	DB 2;	Length 1404;	
Best Local Similarity		96.9%;	Pred. No. 5.6e-143;			
Matches 1361;		Conservative	0;	Mismatches	0;	Indels 43; Gaps 1;
QY	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60			
Db	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60			
QY	61	KRVCTAELSCKRCFCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	120			
Db	61	KRVCTAELSCKRCFCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	120			
QY	121	PPSGASQTIKSTTKRSPKPNKKTKVIBSEITE-----	156			
Db	121	PPSGASQTIKSTTKRSPKPNKKTKVIBSEITE-----	156			
QY	157	-----VKDNKQNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDST	240			
Db	181	KIKSSKNSAANRELQKLVKDNKQNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDST	240			
QY	198	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQTSIDG	257			
Db	241	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQTSIDG	300			
QY	258	KEXTTSAKETQSIKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPTKBPAS	317			
Db	301	KEXTTSAKETQSIKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPTKBPAS	360			
QY	318	TTPKETPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP	377			
Db	361	TTPKETPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP	420			
QY	378	APTTKSAPTTPKBPAPTTPKKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK	437			
Db	421	APTTKSAPTTPKBPAPTTPKKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK	480			
QY	438	EPAPTAKKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTTKBPAP	497			
Db	481	EPAPTAKKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTTKBPAP	540			
QY	498	TTKSAPTTPKBPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTTKBPAPTTTKP	557			
Db	541	TTKSAPTTPKBPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTTKBPAPTTTKP	600			
QY	558	APTAPKBPAPTTPKETAPTPKKLTTPTPKELAPTTPEKBPAPTTPELAPTTPEPTPT	617			
Db	601	APTAPKBPAPTTPKETAPTPKKLTTPTPKELAPTTPEKBPAPTTPELAPTTPEPTPT	660			
QY	618	PEBPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKETAPTPKGTAPTTLKEP	677			
Db	661	PEBPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKETAPTPKGTAPTTLKEP	720			
QY	678	APTTPKKBPAPTTKEPTSTSDKAPTTTPKGTAPTTPKBPAPTTPKBPAPTTPKBPAPTPKG	737			
Db	721	APTTPKKBPAPTTKEPTSTSDKAPTTTPKGTAPTTPKBPAPTTPKBPAPTTPKBPAPTPKG	780			
QY	738	TAPTLKEBPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKGTAPTTPKBPAPTTPKBP	797			
Db	781	TAPTLKEBPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKGTAPTTPKBPAPTTPKBP	840			
QY	798	KPAPTTPETPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENPKBPAGVPT	857			
Db	841	KPAPTTPETPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENPKBPAGVPT	900			
QY	858	TKTPAATKPEMTTAKQKTBDRJLTPPETTTAAPKMTKETATTTTEKTTESKITATTQV	917			
Db	901	TKTPAATKPEMTTAKQKTBDRJLTPPETTTAAPKMTKETATTTTEKTTESKITATTQV	960			
QY	918	TSTTTQDTPPKITLTKTTLAPKVTITTKITITTEINNKPEETAKPKDRATNSKATTPK	977			
Db	961	TSTTTQDTPPKITLTKTTLAPKVTITTKITITTEINNKPEETAKPKDRATNSKATTPK	1020			
QY	978	PQKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN	1037			

Db	1021	PQKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN	1080			
QY	1038	QTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVNOGIIINPMLS	1097			
Db	1081	QTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVNOGIIINPMLS	1140			
QY	1098	DETNICNGKPVDLGTLTTRNGTLVAFRGHYFWMLSPEFSPARRITEVWGIPSPIDTFT	1157			
Db	1141	DETNICNGKPVDLGTLTTRNGTLVAFRGHYFWMLSPEFSPARRITEVWGIPSPIDTFT	1200			
QY	1158	RCNCEGKTPFFKDSQWRFNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPSVY	1217			
Db	1201	RCNCEGKTPFFKDSQWRFNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPSVY	1260			
QY	1218	FFKRGSSIOQYIYKQBPVQKCPGRPALNVPVYGMTQVRRRPFERAIGFSQHTTIRIQY	1277			
Db	1261	FFKRGSSIOQYIYKQBPVQKCPGRPALNVPVYGMTQVRRRPFERAIGFSQHTTIRIQY	1320			
QY	1278	SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYYNIDV	1337			
Db	1321	SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYYNIDV	1380			
QY	1338	PSRTARAITTRSGQTLISKVWYNCP	1361			
Db	1381	PSRTARAITTRSGQTLISKVWYNCP	1404			
RESULT 2						
Q9BX49						
ID	Q9BX49	PRELIMINARY;	PRET;	1404	AA.	
AC	Q9BX49;					
DT	01-JUN-2001	(T-EMBLrel. 17, Created)				
DT	01-JUN-2001	(T-EMBLrel. 17, Last sequence update)				
DT	01-MAR-2004	(T-EMBLrel. 26, Last annotation update)				
DE	BG174L6.2	(MSF: megakaryocyte stimulating factor)				
GN	Name=BG174L6.2;					
OS	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Wray P.;					
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AL133553; CAC36090.1; -					
DR	HSPF; P04004; IOC0.					
DR	InterPro; IPR000585; Hemopexin.					
DR	InterPro; IPR001212; Somatomedin_B.					
DR	Pfam; PF00045; Hemopexin; 2.					
DR	Pfam; PF01033; Somatomedin_B; 2.					
DR	PRINTS; PR00022; SOMATOMEDINB.					
DR	SMART; SM00120; HX; 2.					
DR	SMART; SM00201; SO; 2.					
DR	PROSITE; PS00024; HMOPEXIN; UNKNOWN 1.					
DR	PROSITE; PS00524; SOMATOMEDIN B; 2.					
SQ	SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;					
Query Match						
Best Local Similarity		99.6%;	Score 7293.7;	DB 2;	Length 1404;	
Matches 1358;		Conservative	0;	Mismatches	3;	Indels 43; Gaps 1;
QY	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60			
Db	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60			
QY	61	KRVCTAELSCKRCFCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	120			
Db	61	KRVCTAELSCKRCFCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	120			
QY	121	PPSGASQTIKSTTKRSPKPNKKTKVIBSEITE-----	156			
Db	121	PPSGASQTIKSTTKRSPKPNKKTKVIBSEITE-----	156			


```

QY 157 -----VKDNKKNRTKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 197
Db 181 KIKSSKNSAANRELQKKLVKDNKKNRTKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240
QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKTETVETKETTNNKOTSTDG 257
Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKTETVETKETTNNKOTSTDG 300
QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTPKBPPTTPKBPAS 317
Db 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTPKBPPTTPKBPAS 360
QY 318 TTPKEPTTTIKSAPTTTPKBPAPTTTKGAPTTTPKBPAPTTTKGAPTTTPKBPAPTTTKP 377
Db 361 TTPKEPTTTIKSAPTTTPKBPAPTTTKGAPTTTPKBPAPTTTKGAPTTTPKBPAPTTTKP 420
QY 378 APITTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 437
Db 421 APITTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 480
QY 438 EPAPTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 497
Db 481 EPAPTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 540
QY 498 TTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 557
Db 541 TTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 600
QY 558 APITAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 617
Db 601 APITAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 660
QY 618 PEEBPAPTTPKAAADNTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 677
Db 661 PEEBPAPTTPKAAADNTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 720
QY 678 APITPKBPAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 737
Db 721 APITPKBPAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 780
QY 738 TAPITLKEBPAPTTPKBPAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 797
Db 781 TAPITLKEBPAPTTPKBPAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 840
QY 798 KPAPTTTPETPPPTTSEVSTPTTTKEPTTIHKSDESTPELSABPTPKALENSKBPQVPT 857
Db 841 KPAPTTTPETPPPTTSEVSTPTTTKEPTTIHKSDESTPELSABPTPKALENSKBPQVPT 900
QY 858 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKWKETATTTTETKTTESKITATTTOV 917
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKWKETATTTTETKTTESKITATTTOV 960
QY 918 TSTTTTODTTPKETTTLTKTTTLAPKVTTKKTIITTEIMNKPEBTAKPDRATNSKATTPK 977
Db 961 TSTTTTODTTPKETTTLTKTTTLAPKVTTKKTIITTEIMNKPEBTAKPDRATNSKATTPK 1020
QY 978 PQKPTKAPKKPTSTKPKTMPRVRPKTTPTRPKMTSTMPELNPTSRIAEMLQTTTRN 1037
Db 1021 PQKPTKAPKKPTSTKPKTMPRVRPKTTPTRPKMTSTMPELNPTSRIAEMLQTTTRN 1080
QY 1038 QTPNSKLVBNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1097
Db 1081 QTPNSKLVBNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
QY 1098 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIFSPIDTFT 1157
Db 1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIFSPIDTFT 1200
QY 1158 RCNCEGKTFFFKDSQWRFNTDIKQAGYKPKIFKGGTGTGQIVAAALSTAKYKNWPESVY 1217
Db 1201 RCNCEGKTFFFKDSQWRFNTDIKQAGYKPKIFKGGTGTGQIVAAALSTAKYKNWPESVY 1260

```

RESULT 3

```

Q6ZMZ5
ID Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Sugiyama A., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Masuho Y., Nagai K., Isozaki T., Sugano S., Nagahari K.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAB18580.1; -.
DR InterPro; IPR000595; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 1.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

```

```

Query Match 63.3%; Score 4634.6; DB 2; Length 933;
Best Local Similarity 63.3%; Pred. No. 6.5e-88;
Matches 889; Conservative 0; Mismatches 1; Indels 514; Gaps 3;
QY 1 MAWKTLPIYLLLLLSVFIQQVSSQDLSSCAGRCGEGYSDATCNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFIQQVSSQ----- 25
QY 61 KRVTCTAELSCGRCFESFERGECDCDAOCKKYDKCCPDYEFCAEVHNPTSPBSKKAP 120
Db 26 -----ELSCGRCFESFERGECDCDAOCKKYDKCCPDYEFCAEVHNPTSPBSKKAP 79
QY 121 PPSGASQITKSTTKRSPKPPNKKTKKVIESBEITE----- 156
Db 80 PPSGASQITKSTTKRSPKPPNKKTKKVIESBEITEEHSVSENQESSSSSSSSSTIR 139
QY 157 -----VKDNKKNRTKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 197
Db 140 KIKSSKNSAANRELQKKLVKDNKKNRTKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 199
QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKTETVETKETTNNKOTSTDG 257
Db 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKTETVETKETTNNKOTSTDG 259
QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTPKBPPTTPKBPAS 317

```


QY	498	TTKSAPTTPKEPSEPTTKSBAPTTPKEBAPTTPKQDAPTTPKEBAPTTPKEBAPTTPK	555
Db	371	-----	370
QY	558	APTAPKEBAPTTPKETAPTTPKKLTPTTPEKLAPTTPKEBAPTTPBELAPTTPBEPPTPT	617
Db	371	-----	370
QY	618	PEEPAPTTPKAAAPNTPKBAPTTPKEBAPTTPKEBAPTTPKETAPTTPKGTAPTTLKEP	677
Db	371	-----	370
QY	678	APTTPKBPAPKELAPTTPKPTSTSDKAPPTTPKGTAPTTPKEBAPTTPKEBAPTTPKG	737
Db	371	-----	370
QY	738	TAPTTLKEBAPTTPKBPAPKELAPTTPKGTSTTSKAPPTTPKETAPTTPKEBAPTTPK	797
Db	371	-----	370
QY	798	KBAPTTPBPTTSEVSTPTTTKEPTTIHKSPDBSTPELSAEPKPKALENSPKBGPVT	857
Db	371	-PAPTTPBPTTSEVSTPTTTKEPTTIHKSPDBSTPELSAEPKPKALENSPKBGPVT	429
QY	858	TKTBAATKPEMTTAKDKTTERDLRTPBTTTAAAPKMTKETATTEKTESKITATTQV	917
Db	430	TKTBAATKPEMTTAKDKTTERDLRTPBTTTAAAPKMTKETATTEKTESKITATTQV	489
QY	918	TSSTTTQDTPPKLTTLTKTTLLAPKVTTTKTLITTEIMNKPDEETAKPORATNSKATTPK	977
Db	490	TSSTTTQDTPPKLTTLTKTTLLAPKVTTTKTLITTEIMNKPDEETAKPORATNSKATTPK	549
QY	978	POKPTKAPKPTSTTKPKTMBPVVRPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN	1037
Db	550	POKPTKAPKPTSTTKPKTMBPVVRPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN	609
QY	1038	QTPNSKLVENAPKSDAGAGBETHMILLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1097
Db	610	QTPNSKLVENAPKSDAGAGBETHMILLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	669
QY	1098	DETNI CNCKPVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPPARITEVWGIPSIDTVFT	1157
Db	670	DETNI CNCKPVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPPARITEVWGIPSIDTVFT	729
QY	1158	RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCI VAALSTAKYKNWPESVY	1217
Db	730	RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCI VAALSTAKYKNWPESVY	789
QY	1218	FFKGGSIQQIYYIQEVPQKCPGRRPALNVPVYGEQTQVRRRPFERAIQFSGTHIRIQY	1277
Db	790	FFKGGSIQQIYYIQEVPQKCPGRRPALNVPVYGEQTQVRRRPFERAIQFSGTHIRIQY	849
QY	1278	SPARLAVQDKGVILHNEVKVSIILWGLPNVTSALSLENKPDGYDYAFSKQOYXNIDV	1337
Db	850	SPARLAVQDKGVILHNEVKVSIILWGLPNVTSALSLENKPDGYDYAFSKQOYXNIDV	909
QY	1338	PSRTARAITRSGQTLSKWYNCP	1361
Db	910	PSRTARAITRSGQTLSKWYNCP	933
RESULT 5			
Q9JM99	1D	Q9JM99	PRELIMINARY; PRT; 1054 AA.
AC	AC	Q9JM99;	
DT	DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	DE	Mus musculus.	
GN	GN	Name=Brq4;	
OS	OS	Mus musculus (Mouse).	
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	

Biochem. Biophys. Res. Commun. 254:535-541(1999).

[2]

RL PTTLKEPAPTPPKKPAKELAPTTTKEPTSTTSDEKPAETTPKGTAPTTTKEPAPTTPKEP 730
 RP SEQUENCE FROM N.A.
 RC TISSUE=Articular cartilage;
 RX MEDLINE=99171663; PubMed=10073655;
 RA Schumacher B.L., Hughes C.E., Kuettnner K.E., Caterson B.,
 RA Aydelotte M.B.;
 RT "Immunodetection and partial cDNA sequence of the proteoglycan,
 RT superficial zone protein, synthesized by cells lining synovial
 RT joints.";
 RL J. Orthop. Res. 17:110-120(1999).
 DR EMBL; AF056218; AAD13404.1; -;
 DR InterPro; IPR000585; Hemoexin.
 DR Pfam; PF00045; Hemoexin; 2.
 DR SMART; SMO120; HX; 2.
 DR PROSITE; PS00024; HEMOEXIN; UNKNOWN_1.
 FT NON TER 1
 SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

Query Match 23.4%; Score 1714.7; DB 2; Length 401;
 Best Local Similarity 80.2%; Pred. No. 4.1e-28;
 Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

QY 964 PKDRATNSKATTPKPKPKTKAPKPKTSTKPKKTPRVRKPKTKTTPRK-MTSTMPENLPT 1022
 DB 5 PKGRATNSQVTTTPKPKPKTKAPKPKTSTKPKT-PRVRKPKTKTTPPKTTTSAMPEPTPT 63
 QY 1023 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMYL 1082
 DB 64 S-LPEAMLQTTTRPTTENSEIIDVNSNEDGAAEGEKPHMIFRPVPLTPIVPGTEII 122
 QY 1083 PRVNOGIIINPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSPARRI 1142
 DB 123 VEGSPQGGINPMFSDETNLCNGRPVGLTTLRNGTLVAFRGHYFWMLSFSPSPARRI 182
 QY 1143 TEVWGIPSPIDTVTRCNCEGKTFPFGKDSQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1202
 DB 183 TEVWGIPSPIDTVTRCNCEGKTFPFGKDSQYWRFTNDIKDAGYKPKLISKGFGGLNGKIV 242
 QY 1203 ALSTAKYKNWPSVYFFKRGSGSIQYIYKQEPVQKCPGRRPALNYPVVGEMTQVRRRRF 1261
 DB 243 ALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTGCTGRRPAINYSVYGEAQVRRRF 302
 QY 1263 RAIGPSQTHIRIQYS-PARLAYQDKGLVHNEVKVSLMRGLPNVVTSAISLPIRKPDGY 1321
 DB 303 RAIGPSQVHTIRIHVTPRVYQDKGFLHNEVKVSLMRGLPNVVTSAISLPIRKPDGY 362
 QY 1321 GYDYAFSKDQYVNDVPSRTARAITTRSGQTLISKIWNCP 1361
 DB 363 DYALSKDQYVNDVPSRTARAITTRSGQTLNTWVNC 401

RESULT 7

MUC2_HUMAN STANDARD; PRT; 5179 AA.

ID MUC2_HUMAN
 AC Q02817; Q14878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Mucin 2 precursor (Intestinal mucin 2).
 GN Name=MUC2; Synonyms=SMUC;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE=94132002; PubMed=8300571;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
 RT Identification of the amino terminus and overall sequence similarity
 RT to prepro-von Willebrand factor.";

Biochem. Biophys. Res. Commun. 254:535-541(1999).

[2]

RL PTTLKEPAPTPPKKPAKELAPTTTKEPTSTTSDEKPAETTPKGTAPTTTKEPAPTTPKEP 730
 RP SEQUENCE FROM N.A.
 RC TISSUE=Articular cartilage;
 RX MEDLINE=99171663; PubMed=10073655;
 RA Schumacher B.L., Hughes C.E., Kuettnner K.E., Caterson B.,
 RA Aydelotte M.B.;
 RT "Immunodetection and partial cDNA sequence of the proteoglycan,
 RT superficial zone protein, synthesized by cells lining synovial
 RT joints.";
 RL J. Orthop. Res. 17:110-120(1999).
 DR EMBL; AF056218; AAD13404.1; -;
 DR InterPro; IPR000585; Hemoexin.
 DR Pfam; PF00045; Hemoexin; 2.
 DR SMART; SMO120; HX; 2.
 DR PROSITE; PS00024; HEMOEXIN; UNKNOWN_1.
 FT NON TER 1
 SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

Query Match 23.4%; Score 1714.7; DB 2; Length 401;
 Best Local Similarity 80.2%; Pred. No. 4.1e-28; Indels 3; Gaps 3;
 Matches 320; Conservative 24; Mismatches 52;

QY 964 PKDRATNSKATTPKPKPKTKAPKPKTSTKPKKTPRVRKPKTKTTPRK-MTSTMPENLPT 1022
 DB 5 PKGRATNSQVTTTPKPKPKTKAPKPKTSTKPKT-PRVRKPKTKTTPPKTTTSAMPEPTPT 63
 QY 1023 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMYL 1082
 DB 64 S-LPEAMLQTTTRPTTENSEIIDVNSNEDGAAEGEKPHMIFRPVPLTPIVPGTEII 122
 QY 1083 PRVNOGIIINPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSPARRI 1142
 DB 123 VEGSPQGGINPMFSDETNLCNGRPVGLTLRLNGTLVAFRGHYFWMLTPTTPPPPPRRI 182
 QY 1143 TEVWGIPSPIDTVTRCNCEGKTFPFGKDSQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1202
 DB 183 TEVWGIPSPIDTVTRCNCEGKTFPFGKDSQYWRFTNDIKDAGYKPKLISKGFGGLNGKIVA 242
 QY 1203 ALSTAKYKNWPSVYFFKRGSGSIQYIYKQEPVQKCPGRRPALNYPVVGEMTQVRRRPF 1261
 DB 243 ALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTGCTGRRPAINYSVYGETAQVRRRPF 302
 QY 1263 RAIGPSQTHIRIQYS-PARLAYQDKGLVHNEVKVSLMRGLPNVVTSAISLPIRKPDGY 1321
 DB 303 RAIGPSQVHTIRIHVTPRVYQDKGFLHNEVKVSLMRGLPNVVTSAISLPIRKPDGY 362
 QY 1323 GYDYAFSKDQYVNDVPSRTARAITTRSGQTLISKWYNCP 1361
 DB 363 DYALSKDQYVNDVPSRTARAITTRSGQTLNTWYNCP 401

RESULT 7

MUC2_HUMAN STANDARD; PRT; 5179 AA.

ID MUC2_HUMAN
 AC Q02817; Q14878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Mucin 2 precursor (Intestinal mucin 2).
 GN Name=MUC2; Synonyms=SMUC;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE=94132002; PubMed=8300571;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
 RT Identification of the amino terminus and overall sequence similarity
 RT to prepro-von Willebrand factor.";

	3.	1433	1448	
FT REPEAT		1433	1448	
FT REPEAT		1449	1464	4.
FT REPEAT		1465	1471	5.
FT REPEAT		1472	1478	6.
FT REPEAT		1479	1494	7A.
FT REPEAT		1495	1517	7B.
FT REPEAT		1518	1533	8A.
FT REPEAT		1534	1556	8B.
FT REPEAT		1557	1572	9A.
FT REPEAT		1573	1596	9B.
FT REPEAT		1597	1612	10A.
FT REPEAT		1613	1635	10B.
FT REPEAT		1636	1651	11A.
FT REPEAT		1652	1675	11B.
FT REPEAT		1676	1683	12.
FT REPEAT		1684	1699	13.
FT REPEAT		1700	1715	14.
FT REPEAT		1716	1731	15.
FT REPEAT		1732	1747	16.
FT DOMAIN		4815	4886	VWFC 1.
FT DOMAIN		4924	4991	VWFC 2.
FT DOMAIN		5075	5160	CTCK.
FT DISULFID		5075	5122	By similarity.
FT DISULFID		5089	5136	By similarity.
FT DISULFID		5098	5152	By similarity.
FT DISULFID		5102	5154	By similarity.
FT DISULFID		?	5159	By similarity.
FT CARBOHYD		163	163	N-linked (GlcNAc. .)
FT CARBOHYD		423	423	N-linked (GlcNAc. .)
FT CARBOHYD		670	670	N-linked (GlcNAc. .)
FT CARBOHYD		770	770	N-linked (GlcNAc. .)
FT CARBOHYD		894	894	N-linked (GlcNAc. .)
FT CARBOHYD		1139	1139	N-linked (GlcNAc. .)
FT CARBOHYD		1154	1154	N-linked (GlcNAc. .)
FT CARBOHYD		1215	1215	N-linked (GlcNAc. .)
FT CARBOHYD		1230	1230	N-linked (GlcNAc. .)
FT CARBOHYD		1246	1246	N-linked (GlcNAc. .)
FT CARBOHYD		1787	1787	N-linked (GlcNAc. .)
FT CARBOHYD		1820	1820	N-linked (GlcNAc. .)
FT CARBOHYD		4339	4339	N-linked (GlcNAc. .)
FT CARBOHYD		4351	4351	N-linked (GlcNAc. .)
FT CARBOHYD		4362	4362	N-linked (GlcNAc. .)
FT CARBOHYD		4373	4373	N-linked (GlcNAc. .)
FT CARBOHYD		4422	4422	N-linked (GlcNAc. .)
FT CARBOHYD		4438	4438	N-linked (GlcNAc. .)
FT CARBOHYD		4502	4502	N-linked (GlcNAc. .)
FT CARBOHYD		4616	4616	N-linked (GlcNAc. .)
FT CARBOHYD		4627	4627	N-linked (GlcNAc. .)
FT CARBOHYD		4752	4752	N-linked (GlcNAc. .)
FT CARBOHYD		4787	4787	N-linked (GlcNAc. .)
FT CARBOHYD		4881	4881	N-linked (GlcNAc. .)
FT CARBOHYD		4888	4888	N-linked (GlcNAc. .)
FT CARBOHYD		4955	4955	N-linked (GlcNAc. .)
FT CARBOHYD		4970	4970	N-linked (GlcNAc. .)
FT CARBOHYD		5019	5019	N-linked (GlcNAc. .)
FT CARBOHYD		5038	5038	N-linked (GlcNAc. .)
FT CARBOHYD		5069	5069	N-linked (GlcNAc. .)
FT CONFLICT		1351	1351	H -> L (in Ref. 3).
FT CONFLICT		1412	1412	T -> S (in Ref. 3).
FT CONFLICT		1449	1449	L -> P (in Ref. 3).
FT CONFLICT		1504	1504	M -> T (in Ref. 3).
FT CONFLICT		4192	4192	G -> S (in Ref. 2).
FT SEQUENCE		5179 AA;	540295 MW;	85CD7571FPB9A5663 CRC64;
			19.3%;	Score 1413.6; DB 1; Length 5179;
			Best Local Similarity 19.8%;	Pred. No. 9.5e-20;
			Matches 461; Conservative 90;	Mismatches 518; Indels 1254; Gaps 74;
QY	6	LPFIYLLLSVRFVIQQVSSQDLSSCAGRCG-----EG-		-----YSR 40
		: : : :		:
Db	502	LQVLAPVMQLFVTLDQASQ--GGVQGLCGNFENGLEGGDFKTAGSLIVEATGAGFANTWKA		559
		: : : :		:
OY	41	DATCN-----		---CDY--- 48

QY 1120 VAFRGHYFWMLSPSPPP-----SPARRITEVWGIPSIDT 1154
Db 2712 -----VTPPTPTGTQTPPTTPTTTTPTTPTPTGT 2744

RESULT 8
Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgs1; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yazdell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer G.D., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Brokstein P., Brothman S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foele C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Koira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage D., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RT Science 287:2185-2195(2000).
RT [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426055; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan T.J., Muzny D.M., Nelson C.R.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Fark S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RT Genome Biol. 3:R5EACH0079-R5EACH0079(2002).
RL

QY	1122	FRGHYFWMSPSPSPARRITEVWGIPSPIDVTFRNCBCKTFFFKDSQWRFNDIK	1181
Db	1688	-----PLPLPUP-----NP-PLPLPEI-----	1703
QY	1182	DAGYKPKPIKGGGLTGQVIALSTAKYKNWPSVFFKRGSIQYIYKQBPVKCPGR	1241
Db	1704	-----PLPL-----PPLPTA-----LPLPLPLPL	1723
QY	1242	REALNVPVYGVEMTQVRRRFRFAIGBSQTHIRIQYSPARLAYQDKGLHNEVKVILWR	1301
Db	1724	PP-----	1725
QY	1302	GLENVVTSALSLPNRKP 1320	
Db	1726	LPEVNLTAISLPEISLPN 1743	
RESULT 10			
ID	AAS64673	PRELIMINARY; PRT; 1761 AA.	
AC	AAS64673;		
DT	01-APR-2004 (TrEMBLrel. 27, Created)		
DT	01-APR-2004 (TrEMBLrel. 27, Last sequence update)		
DT	01-APR-2004 (TrEMBLrel. 27, Last annotation update)		
DE	CG33300-PA (Fragment).		
GN	CG33300.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
[1]			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Anatitides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.E.,		
RA	George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., D.,		
RA	Brandon R.C., Rogers Y.H., Blazer R.G., Champagne M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,		
RA	Abriel J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattler P.,		
RA	Burtis K.C., Busem D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Gelbart W.M., Glasser K.,		
RA	Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Harris M.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Laoko P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,		
RA	Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,		
RA	Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster.";		
RL	Science 287:2165-2195(2000).		
[2]			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22426065; PubMed=12537568;		
RA	Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,		
RA	Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,		
RA	George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,		
RA	Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,		
RA	Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,		
RA	Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;		
RT	"Finishing a whole-genome shotgun: release 3 of the Drosophila		
RT	melanogaster euchromatic genome sequence.";		
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).		
[3]			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22426069; PubMed=12537572;		
RA	Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,		
RA	Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,		
RA	Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,		
RA	Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,		
RA	Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,		
RA	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,		
RA	Lewis S.E.;		
RT	"Annotation of the Drosophila melanogaster euchromatic genome: a		
RT	systematic review.";		
RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).		
[4]			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22426070; PubMed=12537573;		
RA	Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,		
RA	Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,		
RA	Ashburner M., Celniker S.E.;		
RT	"The transposable elements of the Drosophila melanogaster euchromatin:		
RT	a genomics perspective.";		
RL	Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).		
[5]			
RP	SEQUENCE FROM N.A.		
RX	FlyBase;		
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.		
[6]			
RP	SEQUENCE FROM N.A.		
RX	FlyBase;		
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AE003626; AAS64673.1; -		
FT	NON TER 1		
SQ	SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;		
Query Match 18.3%; Score 1339.8; DB 2; Length 1761;			
Best Local Similarity 24.1%; Pred. No. 4.1e-19;			
Matches 482; Conservative 106; Mismatches 399; Indels 1012; Gaps 94;			
QY	74	CF-----ESFERGECDCDAQCKYKCCPDYSFCAEV	107
Db	5	CFIWLALYHSANAALGSEIKFGDESTTESAE-----KY-----PDYCW-----	44
QY	108	HNPTSPSSKXKAPPSPSGASQIKTKTKSPFPNNKKTKKVISE-----ITEVKDNKK	162
Db	45	NPELPGC-----PFGEGDGNNTTKISTLATTK-----STVTTSEETTLTKITIKSTAK	94
QY	163	NRTKK---KPTPKPPVDEAGSLDNGFKVTPDTST-----TOHNKV-----	203
Db	95	PTTKTNTNEPTTE-----KITPKATIKSTATTARATAPKTEQTTLR	138
QY	204	-----STSPKITTAKPINRPSLPNSDTSKET---SLTVNKKETTVTKETTTINK---QTSTD	256
Db	139	TTTIKSTSELITTLK-----TTTIKSTAKTSKSTHNTTKSTLTATTEETTRKSTAKT	194
QY	257	GHEKTTSAKEQ-----SIEKTSKADLAPTSKVLAKTPKPAETTTKGPALTTPKEPT--	308
Db	195	TREPTTKKETTERTTQBPSTSKTTTHTHE---TTAEPATKKTTHPTTKQKSTTLRIETPTR	252
QY	309	-----PTTPKEPASTTPKEPT-----PTTIKSA-----	331
Db	253	KSSTAKTTRPTTKETTERTTKETSTKTTTHTTETTABEPATKKTTHPTTKQKSTTLRIETE	312

QY	1225	IQOYIYKQEPVQKCPGRPALNVPYCGMTQVRRRPERAIGPSQTHIRIQYSPARLAY	1284	3118	-----TTPKEP-----	-----TPTTIKSAPTTP-----	335
Db	489	SSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTEST	548	489	SSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTEST	548	
QY	121	IQOYIYKQEPVQKCPGRPALNVPYCGMTQVRRRPERAIGPSQTHIRIQYSPARLAY	180	336	-----KEPA-----PPTTKSAPTTP-----	-----KEPATTTTKEPATTTTKEPAT	371
Db	1285	QDKVLHNEVKVSLIWRGLPNVVTSAISLNPRIKPDGYDYAFSKQYNNIDVPSTARA	1344	549	DTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTEST	608	
QY	181	QDKVLHNEVKVGLIWRGLPNVVTSAISLNPRIKPDGYDYAFSKQYNNIDVPSTARA	240	372	T-----TTKEPA-----PPTTKSAPTTPKEPATTTKPKAPAT-----	-----TPKEP	408
Db	1345	ITTRSGQTLK 1355		609	TESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTEST	667	
QY	241	ITTRSGQTLFK 251		409	AP-----TTPKEPTP-----	-----TTPKEPAT	430
Db				668	STTPGTTTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPE	727	
QY	431	PAP-----TTPKEPAT-----APKEPATTTKEPATTTTKEPAT-----	469	728	PSTTPGTTTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTP	787	
Db	470	ESP-----TTPKEPA-----	490	788	EPSTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSS	847	
QY	491	--TKPEA-----PPTTKSAPTTP-----KEPSTTTKEPATTTTKEPAT-----	527	848	ASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTS	907	
Db	528	-----TTPKKAP-----	543	908	SASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTD	967	
QY	544	-----TTPKEPA-----	554	968	TTMSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTST	1027	
Db	555	K-----KPAPTAP-----KPAPTTPETAATTTKK	580	1028	ESDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPT	108	
QY	581	LTPT-----TPE-----KLAPT-----	595	1088	PTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTR	114	
Db	596	-----KPAPTTPBELAPTTPPEPTP-----	619	1148	TPRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPG	120	
QY	620	-----EPAPTTPKAAAPNTPKAP-----	639	1208	TTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPEST	126	
Db	640	-----TTPKEP-----	655	1268	GTTTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSAST	132	
QY	656	TTKETAPTTPKGAAPTTLK-----PAP-----	694	1328	TITDTTKTTP-----TRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTEST	138	
Db	695	-----TKEPTST-----	722	1386	ASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTSTESTD	144	
QY	723	-----APTTPKE-----	748	1446	TMSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPTST	150	
Db	749	-----TTPKKAPKELA-----	760	1506	STDTMTSSASTPESTPGTTRTTRPTSTESTDTMTSSASTPESTPGTTRTTRPT	155	
QY	761	-----PPTTKGPTSTSDK	77				

```
Db 1566 ESTDTMSSASTPEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPR 1625
QY 775 PA-----PTTPKETAPTKEPAP-----TTPK 797
Db 1626 PTSTESTDTMSSASTPEPSTTGTTRTTPRPTPTDTTMSSASTPEPSTTGTTRTTPR 1685
QY 798 KPAPT-----TPE-----TTPPTT-----SEVSTPTT 820
Db 1686 RPTPTDSTMSSMSSESTPEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPR 1745
QY 821 KEPTTIHKSP-----DESPELGABP-----841
Db 1746 TTPGTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPRPTSTESTDTTMSSAST 1805
QY 842 -----TPKALENSKEP-----GVPTTKPAAKPEMT-----870
Db 1806 PEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPRPTSTESTDTTMSS 1865
QY 871 ---TAKOKTTERDLATP-----EITTAAPKMTKETA 899
Db 1866 SASTPEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPRPTSTESTD 1925
QY 900 TT-----TEKTESKITATTQVSTTQD-----TTPFKIT-----931
Db 1926 TTMSSASTPDSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPRPTSTEST 1985
QY 932 -----TLKUTTLAPKVTTKKTTTTEIMNKPEETAKP-----964
Db 1986 DTMSSASTPEPSTTGTTRTTPR-----TSTERIDTMSSAYTPEPSTTGTTRTTPR 2043
QY 965 ---KDRATNSKATTPKPO-----KPTK-----APKK 987
Db 2044 TSTESTDTMSSASTPEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTRTTPR 2103
QY 988 PTSTK-----KPKTMPVRK-----PKTTPTRK 1011
Db 2104 PTSTESTDTMSSASTPEPSTTGTTRTTPRPTSTESTDTTMSSASTPEPSTTGTTR 2163
QY 1012 MTSTMPRLNPT-GRIAEALQTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHV 1070
Db 2164 TTTPR--TPTDSTMSSMSASTTGTTRTTPRPTSTESTDTTMSSASTPE-----2214
QY 1071 FMEVTPDMOYLPRVFNQGHIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRCHYFWM 1130
Db 2215 --PSTTPG-----TTRTTPRPTD--STMSS-----M 2240
QY 1131 SPSPSPSPA-----RAITEVWGIPIPSIDTPTTRCNCEGKTFPFKDSQYWRFTNDIKDAGY 1185
Db 2241 SSASTPEPSTTGTTRTTPR--RPTPTDIT--MSSSMSSAST 2278
QY 1186 PKPIFKGFGGLT-----GQIVAAALSTAKYKNWPSVYFFKGGSIQO-----1227
Db 2279 PEP--STTPTGTTRTTPRPTDSTMSSMSASTPEPSTTPE-----LLQDDOHL 2327
QY 1228 -----YIYKOE-----PVOKCPGRRPALNPVVGEMTQVRRR 1260
Db 2328 PIPQCLVKCRQPLLRGHRHLVQPELLQDDOHLPIQC--RQPLL-----RSH 2373
QY 1261 FERAIGPSQTHIRIQYSPARLAYQKGVHLNEKVNSILWRGLPNVVT--SAISLPINRK 1318
Db 2374 LRHLVQF-----ELLQDD-----QHLPSDTTMSSASTPE---2403
QY 1319 PDGYDYAFESKQYVNIIDVPSRTARAITRSQT 1352
Db 2404 -----PSTTPG---TTRTTP 2416
```

RESULT 13

Q6SSE6

ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.

AC Q6SSE6;

DT 05-JUL-2004 (TReMBLrel. 27, Created)

```
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
GN Plus agglutinin.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_Taxid=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.; 2003 to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1; -.
DR InterPro; IPR002951; Atrophin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01218; PSTLEXSENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 17.2%; Score 1256.8; DB 2; Length 3409;
Best Local Similarity 14.0%; Pred. No. 7.5e-17;
Matches 398; Conservative 166; Mismatches 460; Indels 1812; Gaps 80;

QY 1 MAWKTLPI-----YLLLLLSV-----16
Db 50 LAWVTIPVDALQFPVNVSVTSSAATDAPSIAIYIGALLDVLVWVWPQPCINDTRRSYD 109
QY 17 -----16
Db 110 APWPSRCALPAVQYGYDEYVIDATVYSGNSLRPFYSTCRYPASSDPYAFLEPYSMWVD 169
QY 17 -----FVIOQVSS-----ODLSSACRCGEGYRD 41
Db 170 DFGIYQFVLGDFDMMWGFVTNVTVDSNLKWDPAWTAQAGMLGGDF-----RD 217
QY 42 ATCN-----CDYNCQHVMCECPDFKRYCTAELSCCKCFESFERGECDCDAOCKKYD-- 94
Db 218 AVWNFGHYCSW--PFVRC-----SSCENDIA 243
QY 95 -----94
Db 244 DPYDPDKIANGIYPAVITALDFRNASLIYYLDFGMFPAGSLDNVWVYLNAYNFIGGLP 303
QY 95 -----KCCPD-----YE-----101
Db 304 ANLPTLLPSLQHLALDHCRAFPDVRGTASLOYQWQVYEGYCSGDGVTGSDGTEY 363
QY 102 -----101
Db 364 VISGMIPEWGDAAVAAAASDSSLPLWANLRTVRLSNQALYGIPEGLRSASSISSWELQ 423
QY 102 -----SFCAEVHN-----109
Db 424 NTELOGLPEFAAPINSLLYLTKLGTHTVWHDPAHTQGGECLAPPSPSPSPRPR 483
QY 110 -----PTSPSSKKAP-----PPSGASOT-----128
Db 484 PEPPLPSPPPPLLPSPVPVPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 543
QY 129 -----IKSTTKRSPKPNKKTKKVIIESEITEVKONKNTKKKPTP 171
Db 544 GICDSFSPMPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 578
QY 172 KPPVVDGAGSLDNGDFKVTTPDT---STQHNKYSTSPKITTAKPIPRPSPSPSPSP 228
Db 579 SPP-----FHPPSPDPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
QY 229 KETSLTVNKETTETKTTTINKQISTDGKENTTSKETQSIETKSADLAPTSLKVLAP 288
Db 621 -----PPSPAPSPSP 629
QY 289 TPKAETTTKGPALTTPKEPTTTPKEPAST-----TPKEPTTTPKSAPTTPKEPAPT 341
```

[illegible]

ID AAS07044 PRELIMINARY; PRT; 3409 AA.
 AC AAS07044;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE Plus agglutinin.
 GN SAG1.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
 RA Goodenough U.W.;
 RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AV450930; AAS07044.1;
 SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

 Query Match 17.2%; Score 1256.8; DB 2; Length 3409;
 Best Local Similarity 14.0%; Pred. No. 7.5e-17;
 Matches 398; Conservative 166; Mismatches 460; Indels 1812; Gaps 80;

 QY 1 MAWKTLPI-----YLLLLLSV----- 16
 Db 50 LAWVTIPVSDALQFVNVSVTSSAATDAPSAIAYIGALLDVLWNVPWQPCINDTRRSYD 109
 QY 17 ----- 16
 Db 110 APWSRCALPAVQGYDYDEYIDATVYSGNSLRPFYSTCRYFAGSDPYAFLEPYMPWVD 169
 QY 17 -----FVIQVSS----- 41
 Db 170 DFGIYQPVLDGDMWGFVNTVDSNLKWDIPANTAQAMVGGQDF-----RD 217
 QY 42 ATCN-----CDYNQHWCECPDKRVCTAELSCKRCFESFERGREGDCDAQCKKYD-- 94
 Db 218 AVNFGTHYCSW---PFVEC-----SSCENYDIA 243
 QY 95 ----- 94
 Db 244 DPYDPDLKIANGIVPAVITALDFRNASLYIYVDFGMFPAGSLDNVWYINLAYNFIGGLP 303
 QY 95 -----KCCPD-----YE----- 101
 Db 304 ANLPTLLPSLQHLALDHCRAITDVRGTASLQYGMQVPTGQPYEYCSGDGVTGSDGTEY 363
 QY 102 ----- 101
 Db 364 VISGMIPEWGDVAVAAASDLSPLWANLRTVRLSNQALYGPPEGLRSASSISSWRLQG 423
 QY 102 -----SFCAEVHN----- 109
 Db 424 NTELCGLPEFAAPINSLLYLTKLGLTHVWHDPAHTQGGELAPPPPPSPSPRPR 483
 QY 110 -----PTSGSSKXAP-----PPSGASOT----- 128
 Db 484 PPPLPPSPPLPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 543
 QY 129 -----IKSTTKRSPKPNKKTKKVISEIEITEVKNKKNRKKKKTP 171
 Db 544 GICDPSMPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 578
 QY 172 KPPVDEAGSLDNGDFKVTTPD-----STQHNKVSITSPKITAKPINRPSLPSPNSDTS 228
 Db 579 SPP-----FHPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
 QY 229 KETSLTVNKETTVEKTTTNTKQSTDGKETTSAKETQSIKTSKOLAPTSKVLAKP 288
 Db 621 -----PPSPAPSPSP----- 629
 QY 289 TPKAETTTKGPALTTPKEPTTPKEPAST-----TPKEPTTPKAPTTPKEPAPT 341

Db 630 APPS-----PAPSPQSPSPVPQPPSPVPSPKPPSPAPSPVPSPSPAPSPSPA 683
 QY 342 TTKSAFTTKKEPAPTTTKKEPAPTTK-----EAPITTKKEPAPTTTKSAFTTKKEPAPTTT 397
 Db 684 PPNAPPPAPLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 743
 QY 398 KKPAPT-----TKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPT- 442
 Db 744 EPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPAP 803
 QY 443 ---APKCAPTTKKEPAPTTTKKEPAPT-----TTKE 470
 Db 804 PPSAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 863
 QY 471 PPTTKKEPAPTTTKSAFTTKKEPAPTTTKSAFTTKKEPAPTTTKSAFTTKKEPAPT- 527
 Db 864 PAPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 922
 QY 528 -----TTPKCAPTTTKKEPAPTTTKKEPAPTTTKKAPTAPK----- 563
 Db 923 SPEPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 982
 QY 564 -----EAPITTKETATTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLT 608
 Db 983 PPSAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1041
 QY 609 -----TPKEPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPT 1041
 Db 1042 PPPPPPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1101
 QY 649 TPK-----EPAPTTTKETATTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTP 674
 Db 1102 PPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1161
 QY 675 -----KEPAPTTTKKAPKELAPTT-----TKETSTTSD 704
 Db 1162 PPTPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1221
 QY 705 KPAPTTKGTAPTTKKEPAPTTKKEPAPTTKGTAPTTKKEPAPTTKGTAPTTKKEPAPTT 1276
 Db 1222 SPAPSPAPSP-APSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1276
 QY 764 TKGPTSTTSKAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 822
 Db 1277 SPAPSPAPSPAPSPSP-APSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1333
 QY 823 PTTIHKSPDESTPESLAPETKALE-----NSPKEGV---PTT 858
 Db 1334 PLPPSPAPLPVPFAGPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1393
 QY 859 KTP-----AATKE----- 867
 Db 1394 PTFPLAPLPDCTLLAQALLSTPDANSVTVSAGLFTSVAPSTPELLASFCTVCSQ 1453
 QY 868 MTFTA----- 872
 Db 1454 LTATLSLVGSSRGNTNGSSGNYNGGDAAIQRIAGNSTDRGSSSGSSGSSW 1513
 QY 873 -----KDKTTERDL-----RTTP----- 885
 Db 1514 GPGTETAEWAVDAVDQGTQYQLSIGGVYTRTVVVDTRTPPSVSGNVTLSANRIKQEPS 1573
 QY 886 -----ETTAAPKMT----- 895
 Db 1574 AVGEASLNALGSKQAMLLTISFSEVPADPDPAASLIIVTGALVAEWAAADKMTFYVLAMT 1633
 QY 896 --KETATTTEKTESKIT----- 911
 Db 1634 LPAELVATAAGSSSSGTSRSGNGTAAAAAAPPAGTTGRRALQOQAAAAAPPPA 1693
 QY 912 ----- 911

Wed Oct 13 12:38:27 2004

seq1-g.rup

```
Db 1694 SGSSSLSGAATANQQQRHVHFLLPATVADAARNGRNDLSLSVELTDNVAASPAVGEA 1753
Qy 912 -ATTTOVTSTTQDTTPFKITLTKLTTTLPKVT----- 943
Db 1754 LAITARTATYP-----AVATITLVAAASSSFAAIRAKGSLLLQGSYHIQMLTMSL 1806
Qy 944 -----TTKKTITTTT----- 953
Db 1807 YLASRGVGREYGEVAFKAVILGVKGNLGPFAAAMPTEKEVTAABQARQVGGLMPIG 1866
Qy 954 ---IMNKPETAATKPKDRATSKATTPKPKP----- 981
Db 1867 NDLLGSSNTTAGSSSGSSSSNSP-PRRPPPPAAGSTGLFESNADASPPPLAVATPA 1925
Qy 982 -----TKAPKKTSTKPKTPRVRKPKTTPTRKWTST--MPELN----- 1020
Db 1926 LPAPLPSTIAAATAAP-----PRLPSP--PPAVGSGSTGLPRLHMQML 1970
Qy 1021 -----PTSRIAEMLQTTTPRNPQTPNSKLVE----- 1046
Db 1971 QPAAAAVAAPPPPPASSSALVLPSPPPPPPPSLLIQASATYVSDMQDLLLYTLVVAAM 2030
Qy 1047 -----VNPKSEDAGGABGETPHMLLRPHVFMPEVTPMDMYLPRVFN 1087
Db 2031 LITAVAAAGRLIAVLYRLLSPE-----PHPLA-----FPRLE- 2064
Qy 1088 QGIIINPMLSDETNICNKPVDGHTLRLNGTLVAFRGH----- 1125
Db 2065 -----TTIAGLLVALTYSCMALGGPAADMHGSRTAAYCV 2100
Qy 1126 -----YFWMLSPSPSPARRITEVWGP-----SPIDTVTRCN- 1160
Db 2101 LTIAVVPVAAFLWLA-----LARAWVPQFTLVEPMTTSSYASPRPSAFERADR 2150
Qy 1161 -----CEGKTFKKDSQY--- 1173
Db 2151 TSPPRSTVSIATAADIGTATCGGGGAAGACADVGAMAAATVAAAGADVPSDDRYARG 2210
Qy 1174 --WR--FTNDIKDAGYKPKIPKFGGLTGQIIVAALSTAKYXNWPESVYFFKGGSIQQYIY 1230
Db 2211 PHWKQF---DGVLPPTVASGVGGSGGVVPL----- 2238
Qy 1231 KOEPVQKCPGRPALNYP-----VYGEWTVQRRREREEA 1264
Db 2239 --PPLVALFGSGRHLPLPLPAAGATAAGGAATAPGSPRGAAGE----- 2284
Qy 1265 IGPSOTHTIRIQYSPARLAYQDKGVHLNEVKVSIILWR----- 1301
Db 2285 ---DTH---QFGP-----H-----WRKFSVPDDGAATAGVGVDAAGGA 2317
Qy 1302 ----- 1301
Db 2318 GGGGGRSGSGSTDGVRPANGTRTPDDGAKGRSSSHGGGNGNAMS SGAASFGASAGG 2377
Qy 1302 -----GLP-----NVVTSALSL----- 1313
Db 2378 GAAAASRLSGRSGSPGDYNAMSGAAMRGAGASRFGPAGSRAGSPPEGAGYAGGA 2437
Qy 1314 -----PNTRKP-- 1319
Db 2438 NAMFAGTAGRAGAISPAHDGSRPFGSGNEGSRQWYSNNAMTAGAVGVGGVRRPAG 2497
Qy 1320 -----DGVDYAFSKDQYNIIDVPS-----RTARAI----- 1345
Db 2498 SLRNASDG-----DAPPLPAPGTVPPLPFANLLTSGADVRSRRLNAEAVA 2543
Qy 1346 -----TTRSG 1350
Db 2544 AAAAGGGGGGTSRSG 2559
```

RESULT 15
Q6SSE8

```
ID Q6SSE8 PRELIMINARY; PRT; 3889 AA.
AC Q6SSE8;
DT 05-JUL-2004 (TREMREL 27, Created)
DT 05-JUL-2004 (TREMREL 27, Last sequence update)
DT 05-JUL-2004 (TREMREL 27, Last annotation update)
DE Minus agglutinin.
GN Name=SADI;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CC-621;
RC Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.,
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450929; AA07042.1; -
DR InterPro; IPR006315; Autotransporter.
DR InterPro; IPR008985; ConA like lec.gl.
DR InterPro; IPR003882; Pistil extensin.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR TIGRFAMs; TIGR01414; autotrans_barli.1.
SQ SEQUENCE 3889 AA; 389223 MW; 4D5B44D5507214A CRC64;

Query Match 16.2%; Score 1184.8; DB 2; Length 3889;
Best Local Similarity 12.1%; Pred. No. 3e-15;
Matches 438; Conservative 187; Mismatches 541; Indels 2452; Gaps 92;

Qy 1 MAWKTLPIYLLLLSVFVIQOVSS-----QQLS----- 28
Db 22 LAFATL-----FPLDLASTQTYGPMNATEDLNEQHGKLLAFILSGDTSFWRSP 71
Qy 29 SCAGRCGG-----YSRATCNCDYNCQHYE-----CC----- 57
Db 72 EVATRLGEGTAPWCISNCQIQTISDOQAPDCESRYCEPFGAALGSENTCCALSLL 131
Qy 58 -----PDFKRVCT- 65
Db 132 DOTYASAQPPSTQAWCSTYPGWGPARGPRPSVCDNFVFAARGTTPAGADDPDLAVSCSS 191
Qy 66 ----- 65
Db 192 GTVPTYISGTRYQRNDTVYRIMHNGAVTNPANVTRNQVSIKLRHSAMWHHPSPVNIISP 251
Qy 66 -----AELSC-----KGRCF----- 75
Db 252 PFFSLVSELACLPLEEIEFEDVALRADYSVLMAQENTTLDNCKVFNLYDPSTYFNVNVL 311
Qy 76 -----ESFERG-----REC-----D 85
Db 312 YNVVFSMLIEPSFMLHSSVSVKASSFNPGALRNAALRTCNFAPFTELHMRWTGQALD 371
Qy 86 CDAQ----- 89
Db 372 WSAEYQAQALSYVRDGSQRLPRGWLPPVIEPVESARLLLPTLRLRTRTRDEQHGAVQT 431
Qy 90 -----CK 91
Db 432 SRPLITGPLGFWALLRNLEYLDLSDMETGALVGPISPTLWMLMHLRVINMTGHEHFCR 491
Qy 92 KYDK-----PD----- 95
Db 492 DWHKIISQIRMYRAATHEPNLVPHYYPGWMGNNGNMTRNYSVYDLSHGHWYDEV 551
Qy 96 -----CC----- 104
Db 552 TTEAGFEVIAHPHGQCCWKWSQTIKDNNEYILYDGSFRFGNNVVQDELYGGFYQDEWC 611
Qy 105 AEVHNPTS-----PPSSKKAP----- 120
Db 612 ----EFTSPQPPPPPPAPPSPSPPTTTPDVPMPSPSPSPAFVMPAPPAPPQPPAPPPLT 667
```



```

QY 121 ----- 120
Db 668 PAAPRRPPLPTWPKWEGAMPFRPPIPRPRPRPPPLPPSPPLPVTPSPRRPPPK 727
QY 121 ----- 120
Db 728 SPFPKPSPPRSPRRPRLPLSPSPPPPPPLPNPSPAPPPPPPPSPSP 787
QY 121 -----PPSGA--SOTIKSTTKRSPKPNKKKKVIRS----- 151
Db 788 PPSPPPPSAPPSAAPPSPDPSPKPPSPVPSPPLPPSPPPSPVPSPPP 847
QY 152 -----BEITEVDKNKNRKKKTPKPPVVVDEAGSLDNGDFKVITPDTSTTOHKNVSTSP 207
Db 848 PASPPTSPAPSPPPPPSPPPSPAPP-----SPPPSPPPSPAPPSP 891
QY 208 KITTAKPINRPSLPNSDTSKETSITVYNKETTIVETKTTTNNKOTSDGKEKTTSAKET 267
Db 892 PLPSPEPPSPAPLPP----- 907
QY 268 OSIETSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTTPKEBPASTTPKEPTTT 327
Db 908 -----PPSPPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 962
QY 328 IKSAPTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 387
Db 963 PLPPPPSPAPSPPPPPSPSPSPAPLPPPPSPSPAPSPPPPPSPSPSP 1022
QY 388 TPKEPAPTTKPPAPTTKPEAP---TTTKEPTTTTKEPAPTTTKEPAPT 443
Db 1023 SPFPSP-EPSPAPSPPPPPSPPPSPAPSPAPSPAPSPAPSPAPSPAPSP 1080
QY 444 PKAPAPTPK--EPAPTTTKEPAP---TTTKEPSTTPKEPAPTTTKEPAPT 497
Db 1081 PPSAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1140
QY 498 TTKSAPTTKEPSPTTTKEPAPTTTKEPAPTTTKEPAPT-----TPKEPAPTTTKEPAP 551
Db 1141 PPSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1198
QY 552 TTTTKEPAPTAKEPAPTTTKEPAPTTPKLTTPTTPEKLAATTEKAPATTPEELAPTTPE 611
Db 1199 HTQSPSP-VPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1255
QY 612 EPTTTPKEPAPTPK-----AAAPNTKEPAPTTTKEPAPTTTKEPAPTTPKETA 667
Db 1256 PFPPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1313
QY 668 GTAPTTTKEPAPTTKPPAPKELAPTTTKEPSTTSDKAPATTTPKGTAPTTTKEPAPTTP 727
Db 1314 -----SPAPSPVPSPSP--APPSPMPPS-----PAPLAPQPPSP-TTPSPAPVP 1355
QY 728 KEAPTTTGTAPTTTKEPAPTTTKEPAPKELAPTTTKEPSTTSDKAPATTTPKETA 787
Db 1356 PSEPPVPVPSPPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1406
QY 788 PKEPAPTPK-----KPAPTTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEP 843
Db 1407 PLQSPFPDPSPQPPSPAPGPPSPSP-----SPSPSTPSPSPAPLAPVPVPMAPQPPS 1460
QY 844 KALENSPKEPGVPTTKTPAATKPE----- 867
Db 1461 PPLPSPPPPPPQPSPTTIPASQPFAPAAVLDGSAATRTSFVASSRSGAFYIAVAVPAS 1520
QY 868 ----- 867
Db 1521 SPSPQVCCSELSYAVLDGASQQVVIPISSGSSSTAGSPVAVTSSVSTPAGAGGLGNG 1580
QY 868 -----MTTAKDKTTERDLR----- 882
Db 1581 THGSTARRALVWEATASSGPAAGVGARHILLATANTSTLEGLLATGRSRAAGMG 1640
QY 883 -----TTPETT----- 888

```

```

Db 1641 MSRLQVVDVQTGCLDPVTAAPPTGTSPNTTSGABGAGSGTVRYSSMGAGSGGLDAAMR 1700
QY 889 ----- 888
Db 1701 LTPGATGDYLLRLKVADQEWWRWVSDIDPPRAAQQLLIARRTTGGSSSSNSTSGSALA 1760
QY 889 ----- 888
Db 1761 AAEDVEQVEVNAHAAGATAVSAQAUVRLMMVIAVSEFPVQPSLTSALKSGARLL 1820
QY 889 -----TAAPKM-----TKETATTEKTTESKITATTTQVITTT-- 922
Db 1821 STQCFASATAAAEVAAGAGTVDASTPPGSDASSAATVAPAAIAFVSGSTSGTSSNSTSG 1880
QY 923 -----QDTTPEKITTLKTTT-----LAPKVTTTKTTTITTEIMN 956
Db 1881 SAYQSCVAVLFAEQDATPELLPPGLTDMHGNINAEPLILSVNLITASADSLSTVERAG 1940
QY 957 KPEETA-----KPKDR 967
Db 1941 APVAAAAGGVFAAFTSASASFLSAFSSRSSLLQSGYHIQMLAMSSSLASPGISPAFR 2000
QY 968 -----ATNSKA----- 973
Db 2001 RISRYLWSLIGIOGNIPLLDGAFSGSAAAGSGSGSSSSSSGGLGDVDVAAVALDRL 2060
QY 974 --TTPKP-----QKP----- 981
Db 2061 QLSVPPPLPAAGDAASQAQPPANLSPSPSASQLVADGSTALAGRRRSRLVQAAAPVADSP 2120
QY 982 --TKAPKKP---TSTKKPKTMR-VRKPKTTP----- 1007
Db 2121 PFTQAPPTAPFGTGTGAAPPAPPPAPPPQSPSTPPPPPPPPALSUGDRDVLVWLOQVGA 2180
QY 1008 -----TPRMTSTMP-----INP----- 1021
Db 2181 IGSNSNSGGGSATSDASVAGYSLGGAASPRGVDVVLDGQLGGLGPGGSAPPMPQPG 2240
QY 1022 -----TSRIAEAMLOTTTTPN----- 1037
Db 2241 SGSGGQSSGASVATDTTTHNVQDLYLTIAIALMLVALVAHAHLVIGLRLAVMYDVC 2300
QY 1038 -----QTPNSKL----- 1045
Db 2301 GAESGVEGLHPVLRFRFAEMVLGGLLLVALTFYSALTLSGAASPRWGDNTAAGRLI 2360
QY 1046 ----- 1045
Db 2361 VLAIVVYGLLLMLLTVCRWYLQEEVDHYMLGPHWQAFDGVIPGGAGAGSGDGGHGA 2420
QY 1046 -----EYNPKSEADAGAGET-----PHMLL----- 1066
Db 2421 SALPAGVFGTGGGGVAFACQAEPPSGDGGGEGDGYGLGPHWALAPAGAKTLDYE 2480
QY 1067 ----- 1066
Db 2481 SIASPAAGAGSAGATRTDAAGSTVGGAGSKGRPRTLRSTDTVTAGGAAAGTAAAAATAA 2540
QY 1067 ----- 1066
Db 2541 AAAAGAPAGAGDVGAFAFORRYGEADNDHGDSDYDGEDGEDLDMAWRRRRPPADGVDA 2600
QY 1067 -----RPHV----- 1070
Db 2601 AAPATAARKMAGSDDGAAGERRRANPHARKSADGAGGAGDPQAGRPFLRRGVSYGDN 2660
QY 1071 -----FMPEVT----- 1076
Db 2661 LLTAGANTRGSALRSTPCPVADDDARSFRFRNTVYAGDNALTSGAASAGAAPSPTRTQ 2720
QY 1077 PDMYDLPRVFNQ----- 1089

```

Wed Oct 13 12:38:27 2004

```

Db      2721 PDSPQAPLSRRGAAMGNAYTAGASRFGGPRTRADSPRAASPPPEPASPLSGDAGALS 2780
Qy      1090 -----III 1092
Db      2781 EVGFNANWVGSRPRSDPTWEDEAAAIROODAADPAVEANRSPKQAGSVEEQOQVAI 2840
Qy      1093 NPMLSDETNI-----CNGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPPS 1137
Db      2841 NPMVPRASVLERPDVAASLAASGAPPQGTVEL-----PP- 2879
Qy      1138 PARRITEVMGIP-----SPIDVTTRCNCCKTFFFKDSQYWFPTNDI----- 1180
Db      2880 -----ITEATG-ENGLGSSPRD-----GRYMTTASVSHGRDTSRLHLHPMARPSLGA 2925
Qy      1181 -----KDAGYKPIFKG----- 1192
Db      2926 SLVPASATNGSGNSGNSGREGSGSTSGGAPAAAMPVRVGGAPSNASAAAAA 2985
Qy      1193 -----FGHITGOIVAALSTAKY-----KNWPE-----SVYFFKGGSIQ----- 1226
Db      2986 AQPAAHQDLFEGELDTTTLTKMTADAGAGAPPENLPSMMRLGSLGLVRRGSGNSGAAQV 3045
Qy      1227 -----QYIY----- 1230
Db      3046 APASPLPAKAPSPPAARPPGAVTTNAAVPMPPQKVRSPSEPANARAQGAAPPAA 3105
Qy      1231 -----KQPPVQKC-----PQRRPAL----- 1245
Db      3106 AAVANGGRELKQOPAVQAVSSITSSIRSNSSGSGNPLARALRYMSVPKGAQPSSTASNA 3165
Qy      1246 ----- 1245
Db      3166 APAIAEAEPAPPSRHTAWGGSLQPPAAAAAGAPPPIAPQSPMRLPMPPPGAGGAGG 3225
Qy      1246 -----NYVYGMTQVRRRFFERAI-----GP----- 1267
Db      3226 AGAKAGGFTSPQASIVMESSPAAGPVWPAARRNRRAIMDEDDLVAEAGGAGPVAE 3285
Qy      1268 -----SQT-----H 1271
Db      3286 AEAPDSAGHYGGWEDEPRLLBREAPSPVMPPPPPLRAGGAARSGSTGMSAAAVSVH 3345
Qy      1272 TIRIQYSPARLAYQDKVLHNEVKVSIWMRG-----LPNVVT----- 1308
Db      3346 S-----YSPSSA-----RGGGSTAGHYMTYNPLSLPSAVSAAAEAAA 3384
Qy      1309 ---SAISLP---NIRKPDGYDYA-----PSKQYNNIDVPSRTARAI- 1345
Db      3385 VTPSALASPLLSLLPGRSQQAAGSPRSVLLILQSRSTFA-----PNPTAMAIL 3435
Qy      1346 -----TTRSGQT 1352
Db      3436 GTRGGGGVGGCSTRSLQS 3453

```

Search completed: October 13, 2004, 12:20:00
Job time : 199.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 27.7222 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKLPYILLLLSVFVIQ.....ARAITTRSGQTLSKVWYNCP 1361

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 79:.*
2: PIR:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1395.9	19.1	3020	2 A439332	mucin 2 precursor,
2	1179	16.1	1664	2 T18262	S-layer protein -
3	1061.2	14.5	7962	2 I38346	elastic titin - hu
4	1050.9	14.4	2187	2 T30826	nascent polypeptid
5	1048.4	14.3	3570	2 T45025	mucin MUC5B, trach
6	1027.3	14.0	1489	2 T31108	cyst germination s
7	1018.4	13.9	1274	2 T16251	hypothetical prote
8	983.3	13.4	3507	2 T34513	hypothetical prote
9	971.4	13.3	6642	2 T29757	protein UNC-89 - C
10	971.3	13.3	1367	1 S48478	glucan 1,4-alpha-g
11	931.3	12.7	4135	2 T46229	tenascin-X - bovin
12	917.8	12.5	1188	2 S49915	extensin-like prot
13	916.8	12.5	5762	2 A18119	proline-rich pepti
14	882.5	12.1	1229	2 T25697	hypothetical prote
15	881.4	12.0	2897	2 B48666	cell proliferation
16	875.7	12.0	1344	1 A35175	mucin 1 precursor,
17	869.3	11.9	3256	2 A48666	cell proliferation
18	862.6	11.8	1151	2 T18535	high molecular mas
19	860	11.7	5262	2 T03454	ALR protein - huma
20	850.6	11.6	3942	2 T42730	Bassoon protein -
21	836.1	11.4	2232	2 T34434	hypothetical prote
22	829.6	11.3	3938	2 T42761	Bassoon protein -
23	828.4	11.3	2142	2 B35098	MHC class III hist
24	823.6	11.2	4006	2 T09070	probable tenascin
25	822.4	11.2	4957	2 T03455	ALR protein - huma
26	820.9	11.2	1832	2 T31113	mucin-like glycopr
27	817.3	11.2	4548	1 S00657	apoptotain(a) (BC
28	809.9	11.1	1872	2 S36132	MHC class III hist
29	802.9	11.0	3381	2 T42389	versican precursor

ALIGNMENTS

RESULT 1

A439332
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU>
A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.O.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33532; MUID:89197956; PMID:2703501
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

30	799.9	10.9	2225	2 T26063	hypothetical prote
31	799.3	10.9	1870	2 S37671	MHC class III hist
32	792.6	10.8	2774	2 A43359	microtubule-associ
33	782	10.7	1611	2 T38236	hypothetical prote
34	777.8	10.6	5105	2 T32650	hypothetical prote
35	777.3	10.6	990	2 I51618	nucleolar phosphop
36	774.4	10.6	761	2 C84672	hypothetical prote
37	773.4	10.6	1630	2 A53577	ascites sialoglyco
38	769	10.5	3566	1 A40701	tenascin-X precurs
39	766.4	10.5	4667	2 T20774	hypothetical prote
40	765.8	10.5	4549	2 T20771	hypothetical prote
41	760.7	10.4	5170	2 T15348	hypothetical prote
42	757.2	10.3	924	2 S27923	gene IF3 protein -
43	756.2	10.3	3968	2 A44265	trithorax homolog
44	755.2	10.3	1777	2 T34369	hypothetical prote
45	754.5	10.3	13055	2 T16580	hypothetical prote

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481; PMID:1985113
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R.Xu, G. Huan, L. Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, R.
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-b
A:Reference number: PQ0328; MUID:92198477; PMID:1550588
A:Accession: PQ0328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M8523
A:Experimental source: small intestine
A:Accession: PQ0329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: lip15.5-lip15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C:Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 19.1%; Score 1395.9; DB 2; Length 3020;
Best Local Similarity 18.7%; Pred. No. 3.8e-21;
Matches 492; Conservative 124; Mismatches 585; Indels 1431; Gaps 90;

Qy 6 LPVLLLLSVFVIQVSSQDLSSCAGRCG-----EG-----YSR 40
Db 502 LQVQLAPVQLFVTLDAQSQ--GQVQLCGNFENGLEDGDFKTAGSLVEATGAGFANTWKA 559
Qy 41 DATCN-----CDV-- 48
Db 560 QSTCHDKLDLWDDPCSLNIESANYAEHWCSSLKLTETPFGRCHSAVDPAEYKRCCKYDTC 619
Qy 49 NCQHYMEC---CPDFKRVCTA-----ELSK 71
Db 620 NQNNEDCLCAALSSYARACTAKVGLWGRHVNCNDVSCPNQVFLYNLTTCQOTCR 679
Qy 72 ---GRCPESF-----ERGR-----ECDC-----DAQCKYDK 95
Db 680 SLSEADSHCLEGAPVDCGCGPDHFLDEKGRCVPLAKSCYHRLGLYLEAGDVVVRQER 739
Qy 96 C----- 96
Db 740 CVCRDGRLHCFQIRLIGQSCFAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVSG 799
Qy 97 ---CPD-----YESFCAEVHNPTSPSSKKAP----- 120
Db 800 CVCPEGLMDGRCGVVEKEPCVHNNDLYSGAKIKYDCNTCTCKGRWVCTQAVCHGT 859
Qy 121 -----PPSG-----ASQT 128
Db 860 CSYSGSHYITFDGKYDFDGHCSYVAVQDYCGQNSILGFSFIITENVPCTGTGVTCSKA 919
Qy 129 IKSTTKRSPKPNKKTKKVIKVESEILTEVDKNKNT----- 165
Db 920 IKIFMR-----TELKLDKHRVVIQRDGHVAVTTREVQYLVV 960
Qy 166 -----KKKPTKPPVVDVDEAGSL-----DNGDFKV----- 190
Db 961 ESSTGIIVWDKRTVTFIKLAPSYKGV-----CGLCGNFDRSNNDFTTRDHMVVSSL 1015
Qy 191 -----TTPDSTTQ-----HNKVTSP----- 207
Db 1016 DFGSNWEKAPTCTPDVSTNPEPCSLNPHRRSMAEKQCSILKSSVFSICHSKVDPKPFYEAC 1075
Qy 208 ----- 207

Db 1076 VHDSCSDTGGDCECFCSAVASYAQECTKEGACVFWRTPDLCPICFYDYNPPHECEWHYE 1135
Qy 208 -----KIITAKPIN----- 216
Db 1136 PCGNRSFETCTRTINGIHNSISVSYLEGCVPRCPKDRPIYEEDLKKCVIADKCGCYVEDTH 1195
Qy 217 --PRPSLP-----PNSDTSKE----- 230
Db 1196 YPGASVPTTEETCKSCVCTNSQVVRPEBGKILNQDGAFCYWEICGPNGTVEKHFN 1255
Qy 231 TSLTVNKET-----TVETKETNTNKOTSTDGKEKITSK----- 265
Db 1256 CSITTRPSLTATFTTITLPTTPTSTFTTTTTTTTTTSSVLSTPKLCCLSWDWINEHPS 1315
Qy 266 -----ETQSIEKTSKAD----- 277
Db 1316 SGSDGDDREFPDGVCAPEDIIECRSVKDPHLSLEHQHGKQVQCDVSFGICKNEQDFGNP 1375
Qy 278 -----LAPTSKVLAKPTPKAETTTKGPAITTPKBPPTTTPKBPAS 317
Db 1376 FGLCYDYKIRVNCWPMDCIITPSPTTTPSPPTTTTTLPTTTPSPPTTTTTPPT 1435
Qy 318 TTPKEPTPTTIKAP--TTPKEPAPTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 376
Db 1436 TTPSPITTTTTLPTTTPSPITTTTTPPTTTPSP--PTTTPSPPTTTPSPPT--TTTT 1493
Qy 377 PAPTPTKAP--TTP-----KEPAPTTKPKAPTTKEPAPTTTKEPAPTTTKEPAPTT 427
Db 1494 PPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPST 1553
Qy 428 TKEPAPTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 487
Db 1554 TTPPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTP 1613
Qy 488 PTTTKEPAPTTTKSAPTTKEPSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 546
Db 1614 PTTTTPSP--PTTTP-----PTSTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPT 1665
Qy 547 KEAPAPTTKAPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 606
Db 1666 SPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSS 1717
Qy 607 PTTPEEPTTTPPEEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 658
Db 1718 TTP--SPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTP 1769
Qy 659 ---KETAPTP-----KG----- 668
Db 1770 FSPFSTTTTTPCPLCNWTLGSLDGKPNFHKGGDTLIGDVCPGWAAISCRATMYP 1829
Qy 669 -----TAPTTLKEPA 678
Db 1830 DVPIGQLGQTVVCDVSGLICKNEQKPGGVIPTMAFCLNVEINVCCECVQPTTM--T 1886
Qy 679 PTTTKEPAPKELAPTT--TKEPTSTSDK-----APT--TPKGTAPTTKEPAPTTTKEPAP 731
Db 1887 TTTTENPTPTTITTTTTVTPTPTTSTQSPNGLOAPTPTTSTTTTTVTPTTPTTPTTPT 1946
Qy 732 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 790
Db 1947 --TPTTTTTTPTPTPTPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2005
Qy 791 PAPTPTKAP-----APTPTPTPTTSEVSTPTTK-----EPTTIHKGDESTPELS 838
Db 2006 PPTTGTQTPMLPSTTTTTVTPTPTTPTTSTGTPPTTSTAPIAELTTSNPPESPSTPQTS 2065
Qy 839 ABPTKALENSKEPGVPTTKTAAKEMTTAKDKTTERDLRTPTTTTAAKMTKET 898
Db 2066 RSTSSPLSTSTLLSTLPPAIEMTSTAPPSTAPTPTTSGGHTLSPPTTTPSPPTTPT 2125
Qy 899 ATTTKTTESKITATTQVTSSTTQDTPFKLTT--LKTTTLAPKVTTKTITTTTTEIM 955

```

Db 2126 GTTT--GSSAAPTPTVQTTTSAWTPPTPLSTPSTIIRTKGLRYPSSVLICCVLNDTY 2183
QY 956 NKPEE-----TAKPKDRATNSKATTPKQKTKAPK 986
Db 2184 YAPGEEVINGTYGDCYFVNCSLCTLBYFNWSCSTPSTPTPSK-STPTPSKSSPTS 2242
QY 987 KPTSTTKKTPRVKRPKTKTPTRKMTSTMPELNPTSRIAEAMLQTTTRPNCOTPSKLVE 1046
Db 2243 KPTPGTKPECFDFDPPR-----QENETWMLCDCFM-ATCKYNNTVIIVKVE 2288
QY 1047 VNP-----KSEDAGGA-----EGTPTPHL 1065
Db 2289 CEPMPMFTCSNGLQPVREDDEGCGWHWECDCYCTGWDGPHYVTPDGLYYSQGNCTVYL 2348
QY 1066 LRPHVFMPEVTPDMYLRPVFNQGIINPMLSDTEINICNGKPVQGLTTLRN----- 1116
Db 2349 VE-----EISPSVD-----NFGYVIDNYHCD-----PNDKVSCTRLLIVRHETQE 2388
QY 1117 -----GTLVAFRG----- 1124
Db 2389 VLIKTVMMPMQVQVQVNRQAVLPYKYGLEVYQSGINVVVDIPELGLVLSYNGLSFSV 2448
QY 1125 ---HYF-----WMLSPFFSP----- 1136
Db 2449 RLPVHRFGNNTKGCGTCTNTSDCILPSGEIVSNCEAAADQWLNDPSKPHCHSST 2508
QY 1137 -----SPARRITEVWGIPSPI-----DTVFTRCN-----CEGKTF 1166
Db 2509 TKRPAVTVGGGKTPHKDCT-----PSPCLQIKDSLFAQCHALVPPQHYVDACVDFSC 2563
QY 1167 FPKDSQV-----WR----- 1175
Db 2564 FMPGSSLECASLOAYALCAQONICLDRNHTHAGCLVECPSHREYQACGPAEPTCKSS 2623
QY 1176 -----FTNDIKDAGY--PKPIFKFGGLTQIIVAAALSTAK 1208
Db 2624 SSQQNNTVLVEGCFCEGTMVYAPGFDVCVKTCGCVGPDNVPRFG----- 2669
QY 1209 YKNWPSVYFFK-----RGGS--IQQYTYKQEPVQKC----- 1238
Db 2670 -----EHFEFDCKNCVLEGGSLICQPKCSQKPVTHCVEDGTYLATEVNPADTCC 2721
QY 1239 -----PGR-----RPNALNP 1248
Db 2722 NITVCKNTSLCKBPSVCPGLFEVSKMVPGRCCPFYWCESKGVGVHGAEBQV--GSP 2779
QY 1249 VYGMTQ----- 1257
Db 2780 VYSSKQDCVCTDKVDNNTLLNVIACTHVPNCNTSCSPGFELMEAPGECCKCEQTHCIK 2839
QY 1258 RRRFERAIGPSOHTIRIQSPARLAYQDKG-----VLHNEVKVSLWGLRGLPNVNTS 1309
Db 2840 R-----PDNQHV--LKPGDFKSDPKNNCTFPSCVKIHNQLISSV-----S 2878
QY 1310 ATSLPN-----IRKPGDYDYAFSKQYQYINIDVPSRTARAITR----- 1348
Db 2879 NITCFNFASICPSITF-----MPNGCCKTCTPRNETRVPCTVPTVTEVS 2926
QY 1349 -----SGQTLISKVWY-----NCP 1361
Db 2927 YAGCTKTVMNHCSGCTFVWYSAKAALDHSCCKEKTQSREVLVLSCP 2978

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J. P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.

```

```

A:Reference number: Z18847; MUID:932099311; PMID:8458832
A:Accession: T18262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUJ>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA4784

```

```

Query Match 16.1%; Score 1179; DB 2; Length 1664;
Best Local Similarity 20.9%; Pred No. 4.4e-17;
Matches 412; Conservative 138; Mismatches 460; Indels 960; Gaps 85;

```

```

QY 4 KTLPIYLLILLISVFVIOQVSSQDLSSCAGRCGEGYSDATCNCDYNCHVMECCP----- 58
Db 6 KVLISILLTLL-----IISTTSWNMS-----FAEATPSIEMV 37
QY 59 -----DFKRVCTAELSCCKGCFESFERGREGDCDCAQCKYDKKCCPD 99
Db 38 LDKTEVHVGVITATIKVNNIRKLAGYQLNIK----- 69
QY 100 YESFCAEVHNPSTP-----PSSKAPPSPGASQITKS----- 131
Db 70 ---FDEVLQPVDPATGEEFTDKSMVYNRVLLTNSKYGPTPV-AGNDIKSGIINFATGYN 125
QY 132 --TKRSP-----KPPNKKTK----- 146
Db 126 NLTAAYSSGIDEHTGIIGEIFKVLKKQNTSIRFEDTILSMGATISGTLSPDWDAAETITGY 185
QY 147 -----KVIESBEI-----TEVK-----DNKN-----RTKKKPT- 170
Db 186 EVIQPDLIVVEABPLKDAVALELDKTKVKVGDIITATIKIENMKNFAGYQLNIKYPDTM 245
QY 171 -----PKPPVVDEAGSGLN----- 185
Db 246 LEAIELETGSAIAKRTWPVTGVLQSDNYGKTTAVANDVGAGIINFAEAYSNTKYRET 305
QY 186 -----GDFKVTTPDT-----STIQNKV 203
Db 306 GVAEBTGIIKGIFRVLKAGSTAIRFEDTTAMPGAIEGTYMWDYGENIKGYSVVQGEI 365
QY 204 -----STSPKLTAKPINRPSLP-----PNS-----DTSK----- 229
Db 366 VAEGEPEGEETPEPVTPTPTPTVTEEPVPSLPSYVIMELDKTKVKVGDIITAT 425
QY 230 -----ETSLTVNKETTETKETTITNTKQSTDGCKET 261
Db 426 IKIENMKNFAGYQLNIKYPDTMLEAIELETGSAIAKRTWPVTGGV-----LQSDNYGKT 480
QY 262 TS-----AKETQSIKTSKDLAPTSKVLAKPTPKAET 294
Db 481 TAVANDVGAGIINFAEAYSNTKYRETGVAEETGIIGKIFRVLKAGSTAI-----RFD 535
QY 295 TT-----KGPALTTP-----KEPTPT-TPKEPASTTPKEP 323
Db 536 TTAMPGAIEGTYMWDYGENIKGYSVQVGEIVAEGBEETPEPVTPTPTVTEEP 595
QY 324 TPT----- 326
Db 596 VPSELPSYVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDKPVLAEFNIETGD 655
QY 327 ----- 326
Db 656 PIDGTPWAVGTTILKNRDYLTGTVAINNVSIGILNFAAYVYVDDYRECKSEDGTGIIG 715
QY 327 -----TIKSAPTPK-----EPATTTKGAPTTPK 351
Db 716 NIGFRVLKAEETTRFEELSMPSGIDGYMLDYNLRISGVYVQIPAKAAS-----D 770
QY 352 EPATTTKGAPTTPKEPA-----PTTTKEPAPTTTKGAPTTTPKEPAPTTKPKPAPTTPK 407
Db 771 EPIPTDTPDEPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSD 830
QY 408 PAPT-----TPK-EPPTPT-TPKEPAPT-TPKEPAPTTPKAPKPA--PTTPK 453

```

Db 831 PTPSDEPTSPDEPTSPDEPTPEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPD 890
Qy 454 EPAPT-TPKEPAPTTTKPSPPTPKPAPTTTKSAPTPTTKPAPTTTKSAPTPTTKPSPPT 512
Db 891 EPTSETPEEPIPTDTPSDEPTSPDEPTSPD-----DEPTSPDEPTSPDEPTSPDEPTSP 947
Qy 513 TKKEPAPTPKKEPAPTPKKA-----PTPKPAPT-TPKEPAPTTTKP 557
Db 948 DTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1007
Qy 558 APTAKPAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTPK 614
Db 1008 EPTSPDEPTSPD-----EPTSPDEPTSPDEPTSPD-----EPTSPDEPTSPD 1046
Qy 615 PT-TPKEPAPTPKAAAPNTPKPAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTP 670
Db 1047 PSETPEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1106
Qy 671 PTTKEPAPTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 729
Db 1107 PTPSDEPTSPD-----DEPTSPD-EPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1161
Qy 730 PAPTTPKAPTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 785
Db 1162 PTPS-----DEPTSPD-----IPTDTPSDEPTSPDEPTSPD-----EPTSPD 1211
Qy 786 TPKPAPTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 844
Db 1212 TPDSPDEPTSPDEPTSPDEPTSPD-----EPTSPD-----PSDEPTP-SDPTP 1262
Qy 845 ALENSPKPAPTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 904
Db 1263 --SETPPEP-IPTDTPSDEPTSPDEPTSPD-----EPTSPD-----PSDEPTP 1319
Qy 905 TTESKATATQVTSITQVTSITQVTSITQVTSITQVTSITQVTSITQVTSITQVTSIT 964
Db 1320 TPEEPIPTDTPSDEPTSPD-----PSDEPTP 1354
Qy 965 KDRATNSKATTPKQ-KPTKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 1014
Db 1355 SDEPTSPD-----TPDELAPTSPRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAE 1401
Qy 1015 -----TPDELAPTSPRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAE 1060
Db 1402 GSGGGGGGGGGTVPSTPTPTS-----KPTSTAPTRE-EPTSPDVPGAIG 1449
Qy 1061 TPEMLLRPHVPMPEVTPMDVLPVPNGIILNPMLSDETNCNGKPDGLTLRLNGTLV 1120
Db 1450 EHAYLR-----GYP-DG----- 1461
Qy 1121 AFRGHYFWMLSPFSPSPAPRITVWVGIPSPIDVTRCNCBGTFFPK-----DSQY 1173
Db 1462 SFR-----PERNIT-----RAEAAVIFAKLLGADSEYGAQ 1491
Qy 1174 -----WRTNDIKDA-----GYPKPIFKGGGLT----- 1197
Db 1492 SASPYSLADTHAAWA-----IKFATSQGLFKGYPDGTGKPDQNTIRAEFATVULHFLTK 1547
Qy 1198 --GO-IVAALSTAKYN-----WPESVVFVKRGGSI-----QOYIYKQ 1232
Db 1548 VKQEIIMKLAIDISNPKFDDCVCHNAQE--FIEKLSLGVISYCPDGTFFKPNYIKRS 1605
Qy 1233 EPVQCPGRRPALNYPVNGEMTQVRRRRFERAIGSQTHTRIQYSPARLAYQDKVLEN 1292
Db 1606 ESV-----ALINRALER--GP----- 1619
Qy 1293 EVKVSILWGLPNVVTSAISLBNIRKPDGYDYAF-----SKDOYINID 1336
Db 1620 -----LINGAPKL-----FQVNB-----SYWAFGDMGALDHSYIIE 1652

elastic titin - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C.Accession: 138346
R.Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A.Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A.Reference number: A57430; MUID:96026330; PMID:7569978
A.Accession: 138346
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-7962 <RES>
A.Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017426
C.Genetics:
A.Gene: GDB:TTN
A.Cross-references: GDB:127867; OMIM:188840
A.Map position: 2q31-2q31

Query Match 14.5%; Score 1061.2; DB 2; Length 7962;
Best Local Similarity 15.1%; Pred. No. 4.2e-13;
Matches 439; Conservative 129; Mismatches 510; Indels 1828; Gaps 107;
Qy 4 KTLPIVLLLLSVFVIQQVSSQDLSSCAGCGEGYGRDATCNDY----- 48
Db 5317 QTEPVTL-----IKDIENQTVLK-----DNDVFEIDIKINYPEIKLSWVKGT 5359
Qy 49 -----NCO-----HYMECC-----PDKR----- 62
Db 5360 EKLEPSDKFEISIDGRHILRVKNCQLKQGNVRLVCGPHIASAKLTVIEPAWHLQDV 5419
Qy 63 -----VCT----- 68
Db 5420 TLKGGQTCIMTVQFSVNVKSEWFRNGRILKPOGRHKTEVEHKVHLKLTADVRAEDQGQY 5479
Qy 69 SCGRCPESFERGRCDDAQCKYDK-----CCPDYESFCAE----- 106
Db 5480 TCK---YEDLETSAELEIAEAPIQTKRIQNIIVVSEHQSATFECEVSFDDAIVTWYKGT 5536
Qy 107 -----VHN----- 109
Db 5537 ELTESQYNFRNDGRCHYMIHNVTPDDEGVSVIARLEPRGEARSTAEALYLTKEIKLE 5596
Qy 110 -----PTSP-----PSSKKAPPPSGASOTIKSTT 133
Db 5597 LKPPDIPDSRVPIPTMBIRAVPPEIIPVAVPVPVLLPTPEEK-PPPKRIEVTKKAV- 5654
Qy 134 KRSPKPNKTKTKVIES-----EETEVKONKNKTKKTP-----KPPVYDE 178
Db 5655 -----KKDAKKVAKPKEMTPREIV-----KKPPPTTLIPAKAPEIIDV 5695
Qy 179 AGSLDNGDFKVTTPDSTTQHKNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKE 238
Db 5696 -----SSKAEVIMT-----ITRKK-----EVOKE-----KE 5718
Qy 239 TTVTKETTTNKT----- 253
Db 5719 AVYEKKQAVHKEKRVFIESFEPEYDELEVEPYTFPEQYVEEDPEDEYIEIKVEAKVEH 5778
Qy 254 -----STDGKETTSAKETOSI----- 277
Db 5779 EWEEDFEGQYVEREGYDEGEBEWEANQEREVIQVQKEVYESHERKVPKAPVEKK 5838
Qy 278 LAPTSKVLAKPT-PAETTT----- 299
Db 5839 APPPPKVPKPVIEKIEKTSRRMEBEKVQVTKVPEVSKKIVPKPSRTFVQEEVTEVKVP 5898
Qy 300 ALTPPK----- 316
Db 5899 AVHTKKMVISBEKMPFASHTSESVTVPEVQKEIVTEEKIHVAISKRVPPKVPPELPE 5958
Qy 317 STTPKEPTTTI-----KSAPTTPKPAPTTTKSAPTTPKBA-----PTTKKBPAP 363
Db 5959 KPAPBEVAVPIPKKVPBPAPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6018

RESULT 4
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 14.4%; Score 1050.9; DB 2; Length 2187;
Best Local Similarity 19.24; Pred. No. 4.3e-14; Gaps 96;
Matches 425; Conservative 161; Mismatches 481; Indels 1151; Gaps 96;

QY	21	QVSSQDLSSCAGRGEGYSRDATCNCYNQHYMECCDFKRVCTAEL-----	68
Db	183	QVFSQGLNLKGT-----PCPPDVVRAPPSHLENPLASVQPLM	222
QY	69	SC-----KGRCFSEFERGECDDCAQCKYDKCCPDY	100
Db	223	SCPQTLNTPVKGVPISALTSRLSLNLKG-----	254
QY	101	ESFCAEVHNTPSPSSKAPP-----	121
Db	255	-----PVSPPARNTAAPSILAPSTSLGCHLPLLHHSSVDSPIQPCQSLAVSNP	305
QY	122	-----PSGASQTIKS	131
Db	306	TSVCHSGTAASCPERCVCVPALPSRLLAVDGAAPSDDKGSSAVTNELCSPGSSNV--A	363
QY	132	TTKRSPPPNKKTKVIESEIEIVDKNKNRTKKP-----	169
Db	364	GTSLSPKASLVPKGSNNVALQVLPASQTKGLKEIPVSCIGATHALDNPSAISVAPA	423
QY	170	TPKPPVDEAGSLDNGDFKVTTPDTS-----TOHNKVSTSP-KITTAKPINRPS	220
Db	424	THVPP-----VSSKDPASPVTSLVVPAAHKQFPAPPASATLGVVSPPLPA	474
QY	221	-----LPPNSDTS-----KE-----TSLTVNKET	239
Db	475	TEGLKNLPIALNVGAPVSPAAGLPTRKDTTLQPLAPIALKESPSSQSASLSLEVLSED	534
QY	240	TVEYKET-----TTTN-----KQSTDGKE	259
Db	535	TVTKTTGGPAPVVRPAIAGVATTTSLRADSPPAVIRADSCVSPNTVSQPLKRSVTDPM	594
QY	260	KITSKET-----QSTIEKT-----	273
Db	595	APRTAKNTAPSTSLVPLASGCPVASSMALSPQNASVSETALALSPEIPKSVFPDPDP	654
QY	274	-----SAKDAPTSKVLAKPT-----PKAETTKGPALT	302
Db	655	LAEISFNARKVDVAVSHMESSGSRQGHDPDASVTAKGTVVCLADSLDTSVSASGSGALS	714
QY	303	TKPEP-----TPTTKPEPASITPEKPTTTIKTSAPTEPK	336
Db	715	GASSPLYPLEVFLPEAGLAVGPKGLNKLSTPTTSPSGKAPVPSTGAPSPKGPVIVT	774
QY	337	E-----PAPTTTKS-----APTTPKE-----	352

Db	775	ESSISSQVPAIBILPSPQKTBEVTASRLISAVQSPKVDPIMSDVTPPTSPPKTSATAVPKD	834
QY	353	-----PAPTTTKKP-APTTPKEPAPTTTKPEAPTTTKS--APTTPKE-----	391
Db	835	TSATLSLXSVPAVTSLSPPKAPVAFSNEA-TIVPTETISLKNALAAATPKEKTLATSIPK	893
QY	392	-----PAP-TTPPK-----PAPTT-----PKE-----PAPTTTPEKT	417
Db	894	VTSPPQKTPKSVSLKGPAMTSKKATEIAASKDVSPSQFPKVEVPLQLHVPTSPPKSPV	953
QY	418	PTT-----PKPAPTTKEPAPTTTKE-PAPTAPKKPAPTTTKEPAP-----TTPKEP	463
Db	954	SDTLGALTSPPPKGP-PATLAETPTYPKSKPAASKK-----TPATPSPEGVAVPLEI	1008
QY	464	APTTPKPSPTTPKEPAPT-TTKSAPTT-TKE-----PAPTTTKGAPTT	505
Db	1009	PCSKKAPKATAPKESASISSKGRAPKTAVSKEIPSGVAVPVLLEISLPLKETSKA--T	1066
QY	506	PKEPSPTTK-BPAPTTTKEPAP-----TTPKKPAP--TTPKE-----PA	542
Db	1067	PGKSSASSPKRSPKTAGPKETPGGVAVPPEISLPPKETPQMATPNESLAASQKSPK	1126
QY	543	PTTPKEPAP-----TTTKKPAPTAP-KEPAPTTPKET-----APTTPKKLTPTT	586
Db	1127	TSVPKETPPGGVTAMPLEIPSAPOKAPKTAVPKQIPTPEDAVTILAGSPLSPKKASATA	1186
QY	587	EKLAPTTPE-----KPAP-----TTPPEELAPTT-----	609
Db	1187	PKAPATPSVGVIAVSSEISPSPKTSKTAAPKENSATLPPKRSKTAAPKETPATSEG	1246
QY	610	-----PEETPTTPEPAPTTPKAA-----APNTPKEPAPTTPKEPAPTTP--K	651
Db	1247	VTAVPSEISPSPTPASKGVFVTLTPKAPNALAESPASPKVKPKTAAPSETSTTPSQK	1306
QY	652	EPAPTTTPEAPT-----TPKGTAPTTLKBP-----PTTPKKPAPKELAPTTT	696
Db	1307	IPKVGAPKEASATPPSKKTPKTAVPKETSAPSEGVAVPVLLEIPSPRK-APKTAAPKETP	1365
QY	697	EPT-----STTSDKPAPTTP-KGT-----APTTP-----	719
Db	1366	APSEGATTAPVQIAPPSPRKSKAGSKETPTTSPSGVTAAPLEIPISSKTSKMASPK	1425
QY	720	-----KEPAPTTTPE-PAPTTPKGAPT	741
Db	1426	ETLVTPSSKKLSQTVGPKET-SLEGATAVPLEIPSHKKAPKTVDPKQVPLTPSK-DAPT	1484
QY	742	TLKEPAPTTPKKAPKELAPTTTKGPTSTTSDDKAPTTPKETAPTTP-KEPAPTTPKPA	800
Db	1485	TLAE-SPSSPKK-APKTAAPPSER-VITVPEKPA--TPKASGTTASKVPVPAETQVA	1539
QY	801	-----PTTPTPTPTTSEVSTPTTKE-----PTTIHKSPPDESTELSEPTPK	844
Db	1540	VSSRETPTVAVPP-----VKNPSSHKTSKTIELKEAPATLPPSPKIPSSKKAPR	1594
QY	845	ALENSPKE-PGVPTTKTPAATKPEMTTAKOTKTTERDLRTTPETT-----	888
Db	1595	T--GAPKEFPASPIK-----VTTSLAQATAPPSSLOKAPSTTIPKENLAAPVLPVS	1644
QY	889	-----TAAPKMTKETATTTTEKTETTESKITATTQVSTTTQDTTTPFKIT	931
Db	1645	SKSPAAPARASASLSPATAAQTAPKEATTIPSCKKAAT-----ETPIETS	1691
QY	932	TLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKAT-----	974
Db	1692	TAPSLGAPK--ETSETSVSKVLMSGPPKKASSKRASTLPTATLPSLKASVLSETATS	1749
QY	975	-----TPKPKPTKAPK-----	987
Db	1750	SGKDSHIGFVSDACSTGTTTPOASEKLPSKGPFTAEMLAAPAPESALAITAPIOKSPG	1809
QY	988	-----PTSTKKPKTTPR--VRKPKTTPTPRMTSTMPENLNTSRIAEMLOT	1032
Db	1810	ANSNSASSPKCPDPSKKDTKGLPSAVALAPQTVPEK-----DTSKAETLLVS	1859

QY 1033 TTRPNQ-----TENSGL 1044
 Db 1860 PAKGSDCLHSPKPGVGSQVATPLAFTSDKVPPEAVSASVAPKAPAPASLTLAPGVPAPL 1919
 QY 1045 VEVNPKSEDAGAEGETHMLRPHVHVEVTPDMYLPVRPNQGIINPMLSDETNICN 1104
 Db 1920 PPKQELLESARGSVLESFSL-----PVAERDEL-----ELIPPEA-VSG 1960
 QY 1105 GKPVDGLTLRNGTLVAFRGHVWMLSPFSP-----PSPARRITEVWGIPSPDVTFT 1157
 Db 1961 GE-----PFQPLVNMPPAPKEA-----GTPAPAPSA-- 1986
 QY 1158 RNCCEGTFPPKDSQYWRFTNDIKDAGYKPIFKFGG-----KQPVKNKSGTSDSDSVPLEEODSTQT 2018
 Db 1987 -----LTGOIVAALSTAKYKNWPESVFFKRGSGIQYIYKQBPVQKCPGRRPALNVPVYGEN 1253
 QY 1196 -----ATQQAQIAAAE-----IDEEPVSRAKQSR----- 2043
 Db 1254 TQVRRRPERAIGPSQTH-----TIRIQYSPARLAYODKGVLHNEVKVSIILWRGLPNVVT 1308
 Db 2044 SEKKARKAMSKGLRQVTGVTIR-----KSKNLF-----VIT 2079
 QY 1309 SAISLPNIRKPDGY-----DYAFSKDQYINIDVPSRTARAITTR-SGOTLSKVWYN 1359
 Db 2080 -----KPDVYKSPASDTYIVFGEAKIEDLSQQAQIAAAEKFVKQGEAVSNIQEN 2128
 RESULT 5
 T45025
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T45025
 R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A:Reference number: Z22899; MUID:97166151; PMID:9013550
 A:Accession: T45025
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-3570 <DES>
 A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A:Experimental source: placenta
 C:Genetics:
 A:Gene: MUC5B
 Query Match 14.3%; Score 1048.4; DB 2; Length 3570;
 Best Local Similarity 13.5%; Pred. No. 1.4e-13;
 Matches 471; Conservative 159; Mismatches 545; Indels 2326; Gaps 107;
 QY 23 SSQDLSSCAGRCG-----EGYS-----RDATC----- 44
 Db 172 SARGTTTCPCROWTEWDEFDYKGEQLGDDVESYDKIRAAAGHLCQOPKDIQCAESPP 231
 QY 45 -----NCD-----YNCQHYMECCPD----- 59
 Db 232 NWTLAQVGKQVHCDVHFLGVCNWEQEGVFKMCMYNRVLVLCSSDDHCRGATTPPPTE 291
 QY 60 ----- 59
 Db 292 LETATTTTQALFSTPQPTSSPGLTRAPPASTAVPTLSEGLTSPTYTLGTATTGQPR 351
 QY 60 ----- 59
 Db 352 QSAGSTFVPGVATSTLPTRSALPGTTGSLGTWRPSQPPLAPTMTATSRARPTGTAST 411
 QY 60 -----FKRVCETAEIS-----CKGR----- 74
 Db 412 ASKEPLTSLAPTLSELSTQAEISTPRTETMTSPLNTTTSQGTTRCQPKCEWTEWFD 471

QY 75 -----FESFER-----GREC-----DCDAQ----- 89
 Db 472 VDFPTSGVAGDMETFERAAGGKMCWAPKASIECAENYPEVSDIQVQLTCSLETGL 531
 QY 90 -CKKYDK-----CCPDYESFC-----AEVHNPTSP----- 114
 Db 532 TCKNEDQGRNMCNPNVNRVLCCDDY-SHCPSTLATSSTATPSSTPGTWTILTKPTTA 590
 QY 115 -----SSKAPP----- 121
 Db 591 TTTASTGSTATASSTOATAGTHVSTTATTPTVTSKATPFSSPGTATATPALRSTATTP 650
 QY 122 -----PSGASQTI----- 129
 Db 651 TATFTAIPTSSSLGTTWTRLSQTTTPMATMTANPSSTPEIVHTSTVLTTATTGATGS 710
 QY 130 ----- 129
 Db 711 VATPSTPGTAHTKVLTITTTTGTFATPSSSPGRTALPVWISTTTTPTTRGSTVTPSSI 770
 QY 130 -----KSTTKRSPKPPNKKTKKVKIESEIEITEVKDNKNR 164
 Db 771 PGTHTTPTVLTTTITVATGSMATPSSSTQTSCTPSSLTIT-----ATTITAGSTTNPS 825
 QY 165 TKKPTPKPPV----- 176
 Db 826 STPGTTPIPPVLTTTATTAAATSTVTPSSALGTHHTPPVNTTATTGHSLSPPSHTV 885
 QY 177 -----DEAGSLDNDGDKVWTPD-----TSTOHNKVSSTPKI-----TTAKP 214
 Db 886 CTAWTSATSGI-LGTHITTEPSTGTSHTPAATTGTTQH-----STPALSSPHSPSRITESP 940
 QY 215 INPRSLPNSDTSKETSITVNKE-----TTV----- 241
 Db 941 PSPGTTTPTGHTTATSRITATATSKRTSTLLPSQTSAPITTVTVNGCEPQCAWSEWLD 1000
 QY 242 ----- 241
 Db 1001 YSYPMPGSGDDFDYTNIRAAAGVACEQPLGLECRAQAQPGVPLRELGOVVECSLDLGL 1060
 QY 242 -----EYKETTITNKQSTDG-----KEKTTISAKE 266
 Db 1061 VCRNRQVQKFKMCFNVEIRVFCNMGHCPSTATSPSTPGTWTILTELTITATT 1120
 QY 267 TQSIKTSKADLAPTSKVLAKPTPKAETTTKG-----PAL----- 301
 Db 1121 TESTGSTA-----TPTSTLRTAPPKVLITATTPTVTSKATPSSSPGTATAPALRSTA 1176
 QY 302 TTP-----KEPTPT-----TPK----- 313
 Db 1177 TPTATSVTPIPPSSSLGTTWTRLSQTTTPTATMTSTATPSSTPETAHTSVLTATATTGA 1236
 QY 314 -----EPAST-----TPKEPTPTI-----KSAPTTPEP-----APT----- 342
 Db 1237 TGSVATPSTPGTAHTTKVPTTTTGTATPSSSPGTALTTPVWISTTTTPTTRGSTVTP 1296
 QY 343 -----TKSAPTTPK----- 351
 Db 1297 SSIPGTHTATVLTITTTTVAIGSMATPSSSTQTSCTPSSLTATTATTATGSTTNPSST 1356
 QY 352 -----BPAP-----TTTKEPAPT-----TPKEPAPTTTKEPAPTTTKS-----APT----- 389
 Db 1357 PGTPIPPVLTTTATTAAATSTVTPSSALGTTHTPPVNTTATTGHSLSFSPSHVTRT 1416
 QY 390 -----KBPAPTTPKKEPAPT-----PKBPAPTTPKPTT-----TT 420
 Db 1417 AWTGATSGTLGTHITBESTGTSHTPAATTGTTQHSPTALSSPHSPSRITESPSPGTT 1476
 QY 421 PKBPAPTTKEPAPTTPK----- 440
 Db 1477 PGHTTATSRITATATPKRTSTILLPSPTSAPITTVTVNGCEPQCAWSEWLDYSYMPMG 1536
 QY 441 PTA-----PKP----- 447

Db	1537	PSGDEFTYSNIRAGGAVCEQPLGLECRAQAGVPLRELQGVVECSLDGLVCRNREQ	1596	Db	2617	HTPPVPNTTATTHGRSLSPSSPHVTRTAWTSATSGTLGTHTEPSTGSHTPAATTGT	2676
Qy	448	-----APTPKEPAPTPKE-----	462	Qy	743	-----LKEPAP-----	791
Db	1597	VGKFMCFNVEIRVFCNHYGHCPSPTASSSTATPSSTPGTTWILTEQTAATTATTGST	1656	Db	2677	TSTFALSSPHSSRTTESPPSP-----GTTTGHATTASRTATATPSKTRTSTLLPSQP	2732
Qy	463	-----PAPTTPKEPSFTPK-----	477	Qy	792	-----	791
Db	1657	ALPSSPTGAPPKVLTSQATPTATSSKATSSSPRTATTLPLVLTSTATKSTATSFTPI	1716	Db	2733	TSAPITVVTTGCEPQCAWSEWLDYSYMPGPGSGDFDYSNIRAGGAVCEQPLGLECR	2792
Qy	478	-----	477	Qy	792	-----APTPPKP	799
Db	1717	PSSTLGTGTSQNRPHPMATMSTIHPSSTPETHTSTVLTATKATTRAISMSPTSTP	1776	Db	2793	ATAQCVPLGELQGVVECSLDGLVCRNREQVKFMCNVEIRVFCNHYGHCPSPTS	2852
Qy	478	-----EPAPTTPKSAPT-----	489	Qy	800	APTTTETPETPT-----SEVSTPTTTTKEPTTHKSPDESTPELSAETPKALENSPKPGVP	856
Db	1777	GTTWILTELTAAATTAALPHGTPSPSTPGTTWILTEPSTATVTVTGSTATASSSTRATA	1836	Db	2853	STAMPSSPTGTTWILTELTATTTASTGSTATP--SSTP--GTAPPPKVLTSPTATPTAT	2909
Qy	490	-----	489	Qy	857	TTKT-----PAATKPEMTTAKDKT-----TERDLRTTPTT--TAA	891
Db	1837	GILKVLTSATPTVVISSRATPSSSPGTATAPALRSTATTPTATSVTAIPSSSLGTAWT	1896	Db	2910	SSKATSSSPRTATTLPLVLTSTATKSTATSVTPIPSSLTGTTGLPEQTTTPVATMSTIH	2969
Qy	490	-----TTKEPAP	496	Qy	892	PKMTKETA--TTTEKITEKITAATTTQVSTTTQDTP-----FKITTLTKTT--	937
Db	1897	RLSQTTPTATMSTATPSSPTETVHTSTVLTATTTRTGVAIPSSPTGTAHTTKVPT	1956	Db	2970	PSSTPETHTSTVLTATK-----ATTRATSSSTPSSPTGTTWILTELTATTTAGTGPTA	3025
Qy	497	TTT-----KSAPTTPKEP-----SPTT-----	513	Qy	938	-----LAPKVTTTKKTTT-----TTEIMNKPEE-----TAKPKDRAAT--	969
Db	1957	TTTGFTATPSSPGTALTPPVWISTVTTPTTRGTSVTPSSIPGTTHTATVLTITTTTAT	2016	Db	3026	TPSSTPGTTWILTELTATTTASTGSTATLSSTPGTTWILTEPSTATVATPSSGATA	3085
Qy	514	-----TKEPAPTPKEPAP-----	536	Qy	970	NSKATTPKEP-----KPTKAPKPTSTKKPYT-----MPVRKPKTTPTRKMT-----	1013
Db	2017	GSMATPSSSTQTSPTPSLTTATTATGTTNPSSTPGTTPIPVLLTTTATPAATSS	2076	Db	3086	SSTQATAGTPHVSTATTPTVTSKATPSSPGTATAPALRSTATTPTATSTAIAPSSS	3145
Qy	537	TPKBPAPTPKEPAPTTT-----	554	Qy	1014	-----STMBELNPTSGIAEAMLOTT-----TRPNQTPNSKL	1044
Db	2077	TVTPSSALGTTHTPPVENTTATTHGRSLPSSPHVTVTAWTSATSGILGTHITEPSTGT	2136	Db	3146	LGTWTRLSQTTPTATMSTATPSSPTETVHTSTVLTITTTATGATSVATPSSPTGTAH	3205
Qy	555	-----KKPAPT-----APKEPAPTP-----KETAPTPKK-----	580	Qy	1045	VEVNPKSEDAG--GAEGETPHMLLRPHVFM-----	1083
Db	2137	SHTPAATGTTQSPALSSPHSSRTTESPPSGTTTGHTRGTSRTATATPSKURTS	2196	Db	3206	TTKVPTTTTGTATPSSSPGTALTPPVWISTTPTTTTSGSTVTPSSIPGTTHTA	3265
Qy	581	LTPTTPEKL-----	589	Qy	1084	RVEN-----QGIINPMLSDETNICNGKPVGLTLTRN-----GTLVAF	1122
Db	2197	TLLPSSPTSAPIITVTTGCEPQCAWSEWLDYSYMPGPGSGDFDYSNIRAGGAVCEQ	2256	Db	3266	RVLTTTATGSMATPSSSTQ-----SGTTPSLTTTATTITATGTTNPSSTPGT-----	3318
Qy	590	-----A	590	Qy	1123	RGHVFWMLSPSP-----	1138
Db	2257	PLGLECRAQAQGVPLRELQGVVECSLDGLVCRNREQVKFMCNVEIRVFCNHYGHC	2316	Db	3319	-----TPIPPVLTSMATTPAATSSKATSSSPRTATTLPLVLTSTATKSTATSFTBIP	3370
Qy	591	PTTPEKAPAPTPPEELAPTT-----	620	Qy	1139	ARRITEVWGIP-----SPIDTVFTRCNCE-----G	1163
Db	2317	PSTATSTATPSSPTGTTWILTKLTATTATTSTGSTATPSSSTQGPAGTPHVSTATT	2376	Db	3371	SSTLWTTWTVPAQTTTPTMSTMTSTHTSTPTTHTSTVLTATMTATNSTATPSSSTLG	3430
Qy	621	PAPTPPKA-----	628	Qy	1164	KTFPFKOSQYWRFTNDIKDAGYKPEIFKGFGLGCIIVAAL-----STAKYKNW	1212
Db	2377	PTVTSKATPSSPGTATAPALRSTATTPTATSTAIPTSSSLGTTWTRLSQTTTPMATM	2436	Db	3431	TT-----RILTELTATTTAATGSTATLSST	3457
Qy	629	-----AAPNTPEKAPT-----TPKEPAPT--TPKEPAPTTPKET--APT	665	Qy	1213	PESVYFFKGGSIQQYLYKQEPVQKCPGRPALNYPVYGMTQVRRRFRFAIGPSQHT	1272
Db	2437	STATPSSPTETVHTSTVLTATTATGATGSAVATPSSPTGTAHTTKVPTTTTGTFTVTPSS	2496	Db	3458	PGTTWI-----LTE-----PSTIAT	3472
Qy	666	PKGTA-----PTT-----	682	Qy	1273	IRIQSPARLAYQDKGVILHNEVKVLSILWRGLPNVVTSAISLDNIRKPDGYDYAFSKQY	1332
Db	2497	SFGTARTPVWISTTTTPTTSGSTVTPSSIPGTTHTPTVLTITTIQPVATGSMATPSSSTQ	2556	Db	3473	VWVPTGSTATSSLTGTAHT-----PKVVTAMATMP-----	3503
Qy	683	KKAPKELAPT-----TTKEPTSTSDKEP-----	726	Qy	1333	YNIDVPSRTARAITTRSGQTL	1353
Db	2557	TSGTTPSLTTATTATGATGTTNPSSTPGTTPIPELTATTATPAATSTVTPSSALGTT	2616	Db	3504	-----TATASTVPSSTV	3516
Qy	727	PKEPAPT-----	742				

cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C;Species: Phytophthora infestans (potato late blight agent)
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T31108
 R;Goernhardt, B.

submitted to the EMBL Data Library, April 1998

A;Reference number: Z20986

A;Accession: T31108

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1489 <GO>

A;Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC7

C;Genetics:

A;Gene: car90

Query Match 14.0%; Score 1027.3; DB 2; Length 1489;
 Best Local Similarity 24.2%; Pred. No. 6e-14;
 Matches 430; Conservative 67; Mismatches 452; Indels 827; Gaps 84;

QY 41 DATCNDYNCOHYMEC-----CPDFKRYCTAELSCKGRCFSEFGRGRCDCDAQ 89
 DQ 32 DATYLSSESG--IPCQGVGAEPVGTACPKAGDVATS-----DCQPY 71
 QY 90 CKKYD--KCPDYBSPCAEVHN-----PTSPSSKKA----- 119
 DQ 72 LLSYNGAVCVAPVDAECALIHDDMWGCEFPKGTGYTSAVEAETIAAYNGESSGWTGHDEV 131
 QY 120 ----- 124
 DQ 132 VQVDEEEETPARVNVYDVTDP:GVNCEVATEPATQGHATEGGKYDTPSTGTQTDYG 191
 QY 125 -----ASQTIKSTTKRSPKPPNKKTKKVIASEP-----ITBVKNKKNRTKKPTPK 172
 DQ 192 NTIHGSTTTEGVTGGYGTDP-----AKVIDGETYLDYPTGITEILED----- 235
 QY 173 PPVDEAGSLDNGDFKVVTPDTSST-----QHNKYSTSPKI- 209
 DQ 236 -----GTTPGYGTGTTDGGTTGGYTTVDNTHETTBEGAGYDAGTREYSITTIVG 286
 QY 210 -----TTAKPIN-----PPSLP----- 222
 DQ 287 YSTEETGQHVGTGYPSEDSDEAPTEGTYVPREETTAAPSEDTTVAPREVTVYAPTEKP 346
 QY 223 -----PNSDT----- 227
 DQ 347 YDVEETTYVTEESTYAPTKSETNAPTERMHALEKPCDEVTMYAPTEETTYAPTEET 406
 QY 228 ---SKETSLTVNKETVE-TKETT-TNKQTSIDGKEKTTSAKETQSIEKTSAKDLAPTS 282
 DQ 407 YAPTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEET-----YAPTE 451
 QY 283 KVLAKPTPKAETTTKGPALTTPKEPTPT-----PKE- 314
 DQ 452 KTYAET-----EETTYAPTEETPYEPTTEETTYAPTKSETTYAPTEETTYAPTEET 508
 QY 315 ----PASTTPKEPTPTTIKSAPTTTPKEPAPT--TTKSAPTTTPKEPAPT--TTKEPAPTTPK 367
 DQ 509 TTYAPAEETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPY 567
 QY 368 EPAPTTKGPAPTTTKSAPT-----TPKEPAPTTPKPAPTTPKEPAPTTPKEPT 417
 DQ 568 EPEETTYAPTEET--YAPTEETMYAPIETTYAPTEETTYAPAEETPYEPTTEETTYAPT 626
 QY 418 PTPKEP-----AP-----TTKEPAPTTPKEP-----APTAPKAPAPTTPKE 454
 DQ 627 EETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEET 686
 QY 455 PAPTTPKEPAPT--TTKEPSPPTPKP-----APT--TTKSAPT-----TTKEP- 494
 DQ 687 YAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIETTYGPT 746
 QY 495 -----APT--TTKSAPT--TPKESPTTTKEP-----APTTPKEPAPTTPKAPAPTTPKE 540

RESULT 7

T16251

hypothetical protein F35A5.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C;Accession: T16251

Db 747 EETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEET 806
 QY 541 PAPT--TPKEPAPTTTKKPAPTAPKEP-----APTPKETAETTPPKKLPT 584
 DQ 807 YAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETTYAPTEET 866
 QY 585 TPEKLAPT--TPKEP-----APTPBELAPT-----TP 610
 DQ 867 EETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETP 926
 QY 611 EEPPTTPKEP-----APTPKAAAPNTPKAPAPT--TPKEPAPTTPKEP-----APT 656
 DQ 927 YEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPT 986
 QY 657 TPKEPAPTTPKGTAP-----TTLKBPAPTTPKKP-----APKE-----LAPT----- 693
 DQ 987 EETTYAPTEETMYAPIETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEET 1046
 QY 694 -----TTKEPTSTTSKPA-----PTTPKGTAPTTPKBPAPTTPKEPAPTTPKGTAPT 741
 DQ 1047 YASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPT 1106
 QY 742 --TLKEPAPTTPKKP-----APKE-----LAPT-----TTKGPTSTTSKPAAPTTP 780
 DQ 1107 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIETTYGPTTEET--YAPTEA 1164
 QY 781 KETAPT--TPKEPAPTTPKKP-----APTPPTPTTSEVSTPT--TTKEPTTIHKSP 830
 DQ 1165 TTVAETETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPT 1222
 QY 831 DESPELSAETPKALENSPKRGVPTTKPAATKPEMTTAKDKTTERDLRITPTETTA 890
 DQ 1223 YEPTTEETTYAPT--EETTYEPTTEETTYAP-----TEETTYAPTEETTY 1263
 QY 891 APKMTKET-----ATT-----TEKTESKITATTQVTSIT--TQDIT--PFKLT-- 932
 DQ 1264 AP--TEETMYAPIDEITYGTEETTYAPTEATTYAPTEETTYAPTEETTYEPTGETTYAPT 1321
 QY 933 LKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTK 992
 DQ 1322 TEETTYAPTEETTYAPM-----EET-----PYEPAEESTSVSTE 1356
 QY 993 KP-KTWPVRVKPKTTPRKMSTMPBELNPTSRIABAMLOTTTRPNQTPNSKLIVENPKS 1051
 DQ 1357 KPCTBEFTDEPTDEPTDE-----PSDEPTD-----EPTDEPTDLPTDE----- 1395
 QY 1052 EDAGGAEGETPHLLRPHVFMPEBVTDPMDYLRVP--NOGIIINPMLSDETNICNGKPD 1109
 DQ 1396 -----PSTPCDNQGI-----NGIGVE 1411
 QY 1110 GLTTLRNGTLVAPRGHYFWMLSPFPSPARRITEVWGI-----PSPIDTVFTRCCEGKT 1165
 DQ 1412 NKVRVYNA-----GIYNTTGP----- 1428
 QY 1166 FFKDQYWRFT-----NDIKDAGYKPKIFKPG--GLTGQIVAAUSTA-----KYKNW 1212
 DQ 1429 ---RNSQSHWSCRCSCYND-----PICHAFSEHQTSDSYCELTSTSDREEDQQNW 1477
 QY 1213 PESVYFFKRGGSQQVIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRRERAIQPSQHT 1272
 DQ 1478 -----LAGNMDR----- 1484
 QY 1273 IRIQYSPARLAYODKGVHLHNEVKVSIILMRGLPNVVV 1308
 DQ 1485 -----NVVT 1488

Db	742	AKPVSLPEPEKKTPVLAKKAPTTPDSEAAADPVSFGFSSKDPKLAKCAPVKPRDPSMKAV	743
QY	684	PKKPAEK-ELAPITTK-----EPTSTTSD-----KPAPTPKGTAPTTP-KEP	722
Db	744	PIKPAKTEVPVAVKKPEPVAKSRDPSPKAKAENSP--VVPPTPVKNPVKKWKPWE	801
QY	723	-----APTPKPA-----PTPKGTAPTTLKEPAPT	750
Db	802	DDDAPEFVNVPEPEKKTPLVAKKTPVKPRDPSFKKAVPAKSTTADAPFVSVKKEPVS	861
QY	751	-PKPAKPE-----LAPITTKGPT-----STTSKDPAPTTPKETAPT-TPKE	790
Db	862	KPKFSPKKAENSPVVPPTPVKNPVKKWKPWEDDDEPTEEVKKSEPEKKTPLVAKKE	921

922 PEKPDAPKVAAKPRDPSKKAVPEKEPAKVAAKPRDLSPKKAIPIPANTQEAPPIPVKN 981

Db 982 PVKKWPPWEDDEP---AEPVSAPEPEKKTPTVLAKKAPAKPRDPSPKKAAPVAAKPD-- 103

1037 -----PKIPEVPTPVKNPVKKWKPWEDDDEPSEVPVSAPEK----- 107

Db 1076 KTVLAKKAPTKEPATKPDSEAAADPVSGTSDPKLS 111

DB 1113 KKAPVEKP-KPTTDPKDDKLKPSPAKKPEKAPPAAPKKWKPVWDDDDPDEFEADFTVPAP 117

QY	1010	KAWISIN----	PELFIKALREPLQI	-----	PAEKTEKPKPEVSKEPPK----	121
Db	1172	SKKPTDTPADP	PLGGPKTKDKLNKKA	-----	PAEKTEKPKPEVSKEPPK----	121

QY 1057 AEGETPHMLLRPHVFMPEVTPDNDYLPRFVENQGIILNPWLSDEIINICNGRFVDGLILKRN 111
Db 1219 -----PTEPPK-----PAAPKK-----WKPP----- 123

Qy	1117	GTLVAFRGHYFWMLSP-----ISPPSPARRRITTEWVGIPSPIDTVFTRCNCCEGKIRFFKD	11
		: : : :	
Db	1235	-----WEDDDPEPEADFTWPAKPKPDTE-----DPADSL-----	1266

QY	1171	SQWRFTNDIKDAGYKP	1188
Db	1264	-----GGKP	1268

RESULT 8
T34513

C;Species: *Caenorhabditis elegans*
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T34513

R. Ravelli, A. J. Vardoulakis, M. A. J. Vardoulakis, submitted to the EMBL Data Library, August 1994
A; Description: The sequence of *C. elegans* cosmid ZK783.
A; Reference number: Z21536

A;ACCESSION: 134513
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Accession: 13507 -FAV-

A; Cross-references: UNIPROT:Q23587; ENBL:U13046; FLDN:AA027413.1; GSDLE:G000000000

A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 605/1; 586/1; 1175/3; 1207/1
3504/1

RESULT 8

T34513
hypothetical protein ZK783.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 23-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34513

R; Ravello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A; description: the sequence of *C. elegans* cosmid ZK783.
A; Reference number: Z21536
A; Accession: T34513
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA

A:Residues: 1-3507 <FAV>
A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PDB:1AAC2418.1; GSPDB:GNC0002
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CBSP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 3504/1

Query Match 13.4%; Score 983.3; DB 2; Length 3507;
 Best Local Similarity 13.2%; Pred. No. 3.3e-12;
 Matches 472; Conservative 206; Mismatches 519; Indels 2377; Gaps 118;

```

QY 7 PYVLLLLSVFVIOQVS-----QDL-----SSC 30
Db 32 PTFVNFDTSTVICOHSSDPTDLHNMSSLCDGKQDCFNVPAMHDEVFPYCEHKQSTC 91
QY 31 AGR-----CGEGYSRDA-----TCNC-----46
Db 92 SGKAGLYDGAPOCYCDSGFSGACELQDKNECLEHPCHNMAQCQNTLGSYECRCCLPGY 151
QY 47 ---DYNQHYMEC-----CPDPKRV-----63
Db 152 EGNHCECTDIDECSDKLTSCPEHSKCNLPPTYNCQTFQPKGNQSGLDKCADINE 211
QY 64 ---CTAELSC-----GRCFESFERGRE-----83
Db 212 CETGAHNCDADEICENSIGSKVCKVNGKSPGYELIDGKCEDVNECGSEKLHKCDVRADCVN 271
QY 84 ---TIGGYECEBEGFDGKNCQKCGHSVKLPKSSCRKNSAICDRHASCHIVLIDCDCKT 331
Db 87 ---DAOCK-----92
Db 332 GYTGDTGITCHDINECAKTPCSDGGRCLNLDGYYVCCNGQDDATCIKQGAFCSGGCG 391
QY 93 ---YDKC-----CPDYESFCAEV-----HNPTS 112
Db 392 DNACSNATCACIDGFRGDPKRCVDINECEVDSVCGVGDRCVNLFGGFKCQHGSTE 451
QY 113 PPSKKA-----PPSGASQTIKSTTKRSPK-----138
Db 452 AECTDAQFSSDSSTISHGADFTTQEQIIEGSGSIQTSSGSLTV--TRGLPKDVELT 509
QY 139 ---PPN-----141
Db 510 TSGRLACTSYCPNSECVGVCBVCYGGNVLVGCEDIDECITEICNIEANWCYNLIG 569
QY 142 ---KTKTKVIESE-----153
Db 570 GFVCCNPTNATHDDCIDFLTKVKIYAMIIIFLLKGLITKKEGLHVIGNEEDTVVATR 629
QY 154 ---ITEVDKNKN-----RTKK-----168
Db 630 SNHSTDQLITVQQSRNFSTGQIILTRGVSSGEAVTQTDDADEFGLAISADLAGSG 699
QY 169 ---PTPKPPVVDEAGSL-----DNQD-----187
Db 690 SGITLPTLLEPKIE--GSGKASGVWVTEDEGEDEDELMEEGSGSWSTTINGTIGSPR 747
QY 188 ---FKVTT-----PDSITT-----198
Db 748 SEGTRVITLLEDGEPEATATKPGISAPDKTGEKSTESDBEKLTVKDGKEAQSOGS 807
QY 199 -----198
Db 808 SATSSGKSEATSGSSSSAKSGTSGSEAGSSGASSSGSGVSGSGSVSTESGFGTS 867
QY 199 -----QHINKVSTSPKI-----TTA 212
Db 868 SSGSVGEATSGTVDSGSPKSSTEKLPTTKNGEKSPISGSDTTGKSSSEETSR 927
QY 213 KPINRPSLPPNS-----DTSKETSITVWK-----237
Db 928 KPIEGSDSITGSGSGGEWFETGSKGHFSGSKSVTSKGPTQSGAEGSGSGPKVPKPG 987
QY 238 ---ETTVETKETTT-----NKQTSIDGKEKTTSAKETQSI 270
Db 988 APEIITDGESSSTGDKSGGKPADKSNKNVPTGDKNPDITDDGEDSTSETSGEQG 1047
QY 271 EKTSADLAP-----TSKVLAKPT-----289

```

```

Db 1048 PKGSKGQPPGDGKSEVKKPTSEVDGPNLSGTGKSNVPLKPTDLP EEGSGILTTSSGG 1107
QY 290 -----PK-----ABTTTKGPAITPKPTPT--TPKE-----PA 316
Db 1108 KNSTFEHGTKLERLPPKTEKSSSETPOLGLEISAGKAPFEDGTSKEVGLIELWESTPG 1167
QY 317 STT-----319
Db 1168 STTILDSVGLIEISGDLTKATKKPHVEIBSGGTDBEITATTTRDVSSTKKPRVEVDGG 1227
QY 320 -----PKPTPTTIKSA-----PTT-----PKP-----338
Db 1228 DNGETSGVDKPTTAPTPSSAESSTSRPTTSEASPEGSGEAGVPESPFDGSGESSTS 1287
QY 339 -----APTTKSATPTPKPAPT-----TKEPAPTTPKAPAPT-----373
Db 1288 APDGVSTSSATAPEVTTASSTFPDAVERSGIPSTSKPTAEPLETT--APSTEVTS 1343
QY 374 -----TKEPAPTTKGAPTTPKAPAPT-----KKPAPTTPKAPAPT-----412
Db 1344 ESGTBEETLPTTEGSGESTTSGAPT--EPATVLPQNRNEKPEPTKDTFALPTTTGA 1401
QY 413 -----PKE-----415
Db 1402 PQANDSSVENTKCTSSDEGLDALCERRTGVCRCEPGFEGAPPKKSCVDVDECATGDHNC 1461
QY 416 -----PT-----417
Db 1462 HESARCONVVGGYACPTGFRKADDDGSCQIDBCTEHNSTCCGANAKCVNKPPTYSCBC 1521
QY 418 -----PTTKPAPTTPK-----429
Db 1522 ENGFLGDGYQCVPTT--KKPCDSTQSSKSHCSNMSCVDTVDSVECKECMGYKKS 1580
QY 430 -----EPAP-----433
Db 1581 VCEIDINECVAKAPCSLINANCNMNGTFCSCQKQYRGVGMCTDINECDRHPCHPAE 1640
QY 434 -----TTPKE-----438
Db 1641 CTNLEGSFKCECHSGFEGDGIKCTNPLERSCEDVEKFCGRVDHVSCLSVRIYNGSLSSV 1700
QY 439 -----PA-----440
Db 1701 CECEPGRFPEKESNSCVDIDECESRNCDPASAVCVNTEGSYRCECAEGYEGGVCTD 1760
QY 441 -----PTAPK-----445
Db 1761 IDECDRGMAGCDSMAMCINRMSCGCKMAGYTGDTGATCIKIEEBPKSDKTACTDEWSRL 1820
QY 446 -----KPAPTTPKAPAP-----457
Db 1821 CELEKKQCTVDEEVPPQCGACLPGHHPINGTCQSLQISGLCAQKNDCKNKAECIDHPS 1880
QY 458 -----T 458
Db 1881 HFCSCPDPGFIGDMICDDVDCECNAMCDDENTKCENTIGSFNCVCLGFKKVDKCVVD 1940
QY 459 TPKEP-----APTTPKEPPTTPKAPPTTKSAPTTKTKEPAPTTPKSAPTTP 506
Db 1941 EKKQNRKELIDEENSSSSNGQKPTTKGTIVSSTATSESTTABPHVTTSISTTST 2000
QY 507 KEPSPTTKPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTAPKEPA 566
Db 2001 KDMTSSKSPENVTMESSEPEVSTSSKSTTASETTIVSSTPSESS-----SSEAPLTSS 2053
QY 567 PTPKETAPTTPKLTPTTPE-----KUAPTPE-----KPAPTTPBELAPTTP 611
Db 2054 PATTTTEV--ITESSVKSTTPKESSSEITVKLSSKSPVETSSSVKSSPSTPS---TTSG 2107
QY 612 EPTPTTPREPAPTTPKAAAPTTPKEP-----APTTPKEPAPTTPKAPPTTPKET---661

```

Db 2108 SVTSTVPETSKSVLSSEAPVTSSTPEVHTSSSTKPSLSASSTTGDTNSTTSTSSLAS 2167
QY 662 --APTTPKGT--APTTLKEPAPTTPKPAPELAPT-----TTKEP 698
Db 2168 VKSATSAPGTSASVAPVKLSLSPDV-SQPSTTKTDFATBESSTVQASSETSGTSVKSTSEP 2226
QY 699 -----TSTSDKAPATTGKTAPATTKEPAPTTPKEPAPTTPKG-----737
Db 2227 ESHVTKLSITSSNPFSSVPVTSKSTPTPEST--EQPTSTPSGSLTPMNSNSELVLT 2284
QY 738 -----TAPTTLKEPAPT-TPKKAPKEL-----APTTPKGT-----768
Db 2285 SEPHVLSLSPDVSSQSTTPNNLSSTVETPKTSSEVSLNSEESTTEAPTTLSPDIL 2344
QY 769 STT-----SDKP-----APTTP-----780
Db 2345 STTTNNLSQSTVSTEDRSBISSENSEKPTSAPELVTSSVTHVASSPDVPTESSEPDLL 2404
QY 781 -----KETAPTP-----KEPAPTTPKP-----APTTP-----805
Db 2405 TGSSTENIPRASSKQTISSPTPTDTTASBEPTKSTMSDPDLSTTSNVLSSESSTIPSSS 2464
QY 806 -----TPPTTSEVSTPTTTKEPT--825
Db 2465 KSPVSSSTEGISVVTSEFKVPVESTISSVLEEDLTKTTPSPILEETTTASETSEPLET 2524
QY 826 -----IHK--SPDESTPELSAEPPTKALENSP-KEPG-----VPTT-----858
Db 2525 SLTVSVRIHELTTSENVPKSESTTTSSSESKSPQBPAGILTSTVVVPTSSVSLITASE 2584
QY 859 -----KTPAATPEM-----TTTAKDKTTERDLRTPTTETTAAPKMTKE 897
Db 2585 IEAITSNTPPKQGRPTITTPKSLVKSTTSFSTVTSSEPSSEKRTTSTVSTVSTTTTEE 2644
QY 898 TATT-----TEKTTESKITATTQVTS-----TTTQDIT 926
Db 2645 TTTSESILLTAAPKPTSTETSEFAPTTAKTSEKPSNVSSPSRKSTENVEFTSQSG 2704
QY 927 PFKITLLKT-----TLAPKVTTKKITTT-----952
Db 2705 SLESSTMSSTSSPETNAPAVTVSSEASTTLENSSTSPTSSEASVKLSLFPESITS 2764
QY 953 -----EIMNKPEE-----960
Db 2765 EAVTVSSRAPAEITMSSEHREISTVSSSESEPEIPLSTTVSPNVVTASSIPSEBILSS 2824
QY 961 -----TAKPKD-----RATNSKA-----TTPKP 978
Db 2825 VTSSSTPRVRLITGTPDDLLIVSVTVPSHGNRRONITASSVPNSSTSPILLPSESLLTTPQ 2884
QY 979 QKPKAKPKPTSTKKPKTMPVRKPK-----TTTPP-----RKMTSTWPEL 1019
Db 2885 PPTTTTAKPATTSGKRGPPSIOPPAEMFTTTPAPPSPNGGYGBETNQEEQVST-----2940
QY 1020 NPTSRIARAMLQTT-----1034
Db 2941 -----TTEAPLSCTVTCHLATCEOSTGVCIORGFIDGTTACSKKSTADCSILPSLAD 2998
QY 1035 ----RPNOTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGI 1091
Db 2999 KAKCDNSTRSCEC-----DAG-----YI-----GDGYV 3021
QY 1092 INPMLSD--ETNLCNGKPV-----DGLTLR-----N 1116
Db 3022 CSPHPQCDVLRDNLCSPEAVQCNRRCOCLPGFTGDGVKCVSIHERASNCQDANAHCVG 3081
QY 1117 GTLVAFRGHYFWMLSPFPSPPARRITVWG-----IPSPIDTV-FT-----1157
Db 3082 GTTCKCNPGYP-----GNGLCCVDDPLDCVHTGICHPNVACNSES 3122
QY 1158 -RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFG-----ITGQIVAAALS 1205
Db 3123 RQCQCS-----SGFSGNGVSCFPKQKSCRTDKSVCA--3152

RESULT 9

T29757

protein UNC-89 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C:Accession: T29757

R:Du, Z.; Le, T.T.; Wilson, R.

A:Submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid C09D1.

A:Reference number: Z20679

A:Accession: T29757

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-6642 <DUZ>

A:Cross-references: EMBL:AF003131; PDB:1A854132.1; GSPDB:GN00019; CESP:unc-89

A:Experimental source: strain Bristol N2; clone C09D1

C:Genetics:

A:Gene: CESP:unc-89

A:Map position: 1

A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 13.3%; Score 971.4; DB 2; Length 6642;

Best Local Similarity 12.8%; Pred. No. 2.3e-11;

Matches 502; Conservative 197; Mismatches 478; Indels 2756; Gaps 132;

QY 34 CGEGYSRDAICNC-----DYNQCH-----52

Db 659 CKEGDVDFCEVEGWPEPELVLDQQLRPSHDFRLQYDGTAKLEIRDAQDDTGVY 718

QY 53 -----YMECCPD-----FKRVCTAB-- 67

Db 719 TVKIQNEFGSIESKAELFVQADPKNHVAFQATIEYEC--DEGEVRFKSVITGDPN 776

QY 68 -----LSCKGRCFSEFERGECDC 86

Db 777 PEIIWFINGKPLSESEKVFISEDGICILITIKDVRHFDGMVTCQGS-----NRLGSASC 831

QY 87 DAQCKKYDKCCPDYESCAEVHNTSPSSKKA-----PPPS-- 123

Db 832 DGRLLK-----VRVPAPPTFNKPLEDKTVQEKSTVFEVDVSGWPEPILT 876

QY 124 -----GAS-----126

Db 877 FTLCGKELKNGEVEIVGHGDFYRISIPNTSMKHDGEIVAKAQNHEGTAEARLTVE 936

QY 127 -----QTIK-----STTKRSPKP-----139

```

Db 937 QEEERSAPFLKDIEDQTVKTEFAVFEITVRGNPNPEVTFWFLNGHKMDQSGPVKIE 996
QY 140 -----PKN-KTKKKVIE-BEBITEV 157
Db 997 AHNHDKHLTIDSAOYAGTVLCRAENAVGRFETKARLVVLAPEKQKKPKPFVEILLVDKET 1056
QY 158 KDNKKNRTKKXPT-----PKPPV----- 175
Db 1057 VDN-----TVVFVRVEGEKPTVTWYLGEBELKQSDRVEIREPDGSIKIRNIK 1107
QY 176 VDEAG-----SGLDNGDFKVT----- 191
Db 1108 IEDAGEIRAVATNSEGSDTKAKLTVOKKPAPFEDLRLPVSLTVEKGSFAVFAHFGIP 1167
QY 192 -----TPDTS-----TQHNKVS----- 204
Db 1168 LPTVWVGKVRDGOEGARVTRDESTVCGASILITIDTATYYEVNHLTISVAENTLG 1227
QY 205 -----TSPKLTITAKPINPRESLPNS 225
Db 1228 AEETGAQLTIEPKKESVVVEKQDLSSEVQXIEIAQOVKEASPEANTT----- 1274
QY 226 DTSKETSITVKNKETTVEKTTT----- 249
Db 1275 -ITMETSLSITSTKTTMTTTEVTSTVGVTVBTEKESSESATTVIGGGSGGVTEGSI VSK 1333
QY 250 -----NKQTSITGCKEKTSAKETQS 269
Db 1334 IEVVKTSQDVRBEGTPKRRVSPAABELPKVEIDSDRKKKSPSPDKKE-----S 1385
QY 270 IEKTSADLAPTSKVLAKPTPKAETTTTKGPALTTPKEPTPTTP-----KEPASTTPKEPT- 324
Db 1386 PEKTEEPASPTKKT-----GDEVKSPKESPASPTKKESKSPAABEVKSPTK 1432
QY 325 -----PTTIKSAPTTP-----KEPAPTTTKSAPTTPKPEAPTTTKEPAPTTTP----- 366
Db 1433 KEKSPSPPTKKESKSPSPKTKTGDEVKEKSP-----PKSP-----TKKESKPEKPEDV 1480
QY 367 -----KEPAPTTT-----KEPAPTTTKSAPT 387
Db 1481 KSPVKKEKSPDATNIVEVSEETIEKETTMTTMTHESESRTSVKKEKTPKVDKPK 1540
QY 388 TP-----KEPAP-TTPKPEAPTTKPEAPTTKPEPTTPKPE-PA 425
Db 1541 SPTKOKSPEKSIETEEKSPVKKEKSPKVEEKPASPTKKESKPEKASPTKKSENVKS 1600
QY 426 PTTKEPAP-----TTPKEPAP-TAPKKPAPTTKPEAPTTTPKPEAPTTTKEPSPTTP 476
Db 1601 PTKKESKSPKSVVEELKSPKESKPEKADDPKSPTKKE-----KSPKESATEDVKSPTK 1654
QY 477 KEAPATTTKSAPTTTKEPAPTTTKSAPT-----TPKESPTTTKE-PAPTTTPKEP 525
Db 1655 KEKSPKEVEKPT-----SPTKKESSPTKKTDDDEVKSPTKKESKSPQVVEKSPASPTKKEK 1709
QY 526 AP-----TTPKPEAPTTKPE-PAPTTTPKEPAP-----TTTKK-----PAPTAPKE 564
Db 1710 SPEKSVVEEVKSPKESKPEKAEKPKSPTKKESKPEKSAAEVKSPTKKESKPEKSAEBK 1769
QY 565 PAPTTPKEATPTPKKL-----TPTTPEKAPTTTPKPEAPTTTPEELAP-----TTPPEPTPT 616
Db 1770 PKSPTKKESSPV---KQADDEVKSPTKKESKPEKVEEKPASPTKKEKTPKESAAEELKSP 1826
QY 617 TPEEPAPTTKAAAPNTPEPAPTTTPKPEAPTTTPKPEAPTTTPKETAPTTTPKGPAPTTLKE 676
Db 1827 TKKESKSPSPKTKTGDESKEKSPKEE-----KPKSP-----TPKSPPGSPK-----KKKS 1874
QY 677 PAPTTPKPKAPK-----ELA----- 691
Db 1875 KSPAEKPPAPKLTDLKLOTVNKTLAHEVUVVEHATECKWFLDGKEITTAOGVTVSKD 1934
QY 692 -----PTTKEPTST----- 701

```

```

Db 1935 DQFERCSIDTTFMGSGTVSVASNAAGSVETKTTELKVLETPKTKKBEFTDKLDRMEVT 1994
QY 702 ----- 701
Db 1995 KGDVTQMDVIALHSPLYKWKYQNGNLLDGNKVITIKNEENKSSLIIPNAQDSGKITVEAS 2054
QY 702 -----TSUKPAPT--PKGTAPTTPKEPAPTTPKE-----PAPT----- 733
Db 2055 NEVGSSSSAQLTVNPPSTTPIVVDGPKSVTIKETETAERKATISGGFPAPTWKVTINEKI 2114
QY 734 -----TPK-- 736
Db 2115 VEESRTITTIKTEDVYTIKISNAKIEQTGVTKVTAQNSAGQDSQADLKVEPNVKAPFK 2174
QY 737 -----GTAPTTLKE 745
Db 2175 SQLTKVADEGEPLRWNLDELDPSPGTEVSWLLNGQPLTKSDTVQVVDHGDGTVHVIAE 2234
QY 746 PAP-----TTPK----- 752
Db 2235 AKPEMSGTLTAKAKNAAGBCETSAKVTVNGNGNKKPEFVQAPQNHETTLEESVKFSIAVTG 2294
QY 753 KPAPK----- 757
Db 2295 KPMENVTWYLNKKLIQSEEVKVKYVHETGKTSIRIOKPLMEHNGTIRVEAENVSGKVOA 2354
QY 758 -----ELAPTTT-----KGPTST 770
Db 2355 TAQLKVDKKTVEPKFTTNDMDROVKEGEDVKFTANVBGPEPSVAWTLNGEPVSKHNIT 2414
QY 771 TSKP-----APTTPKET-----APTTPK----- 789
Db 2415 VTDKGEHTTIEISAVTPEQAGELSCAENPNVSGKKRDVQLAVKVGADPTFAKNLEDELI 2474
QY 790 ----- 789
Db 2475 TEGELTLMDAKLINVRPKPLTWLKDQVEITSDGHYKIVEEDGSLKLSILOKLEDKGR 2534
QY 790 -----EPAPTT 795
Db 2535 ITIKABSEFGVACSASLGVVGRPAKPAFQSDIAPINLTGDTLECKLLITGDPFPFV 2594
QY 796 -----PKK----- 798
Db 2595 KWIGTQVCATEDTEISNANGVYTMKIHGVTADMTGKIKCVAINKAGEVSTEGPLKVA 2654
QY 799 PAPTTPET-----PPPTTS-----EVSTPT 818
Db 2655 PIPVEFETSLCDATCREGDTLKLRAVLGEPPEPVSVYVNGKKLEESQNIKIHSEKGYT 2714
QY 819 TTKEPTT----- 828
Db 2715 VTIKDITCDYSGQWCEAINEYGKATSEATLLVLRGEPDPDFLEWLSNVBEARTGTVVHK 2774
QY 829 -----SPD----- 831
Db 2775 VFTGDPKPSLWYINNKILLNSDLVTIVDDKTSTLTINSFNPDPVHVBIIKCAENDAG 2834
QY 832 -----ESTPELSAE-----PTPKALENSPK-- 851
Db 2835 EVSCTANMITYTSDMFSESESESAQAEFVGGDDLTDESLEEMHRTPTPVW-----APRFT 2891
QY 852 -----EPGVPTTK----- 859
Db 2892 KIKDTKAKXGSHAVPECVVDTKGVCCKWLDKGKBEILIRIVQTRTGPETHITQELVL 2951
QY 860 -----TPAAT-----KPEMTTTAKDKTTERDLRT 893
Db 2952 DNVTPEDAGKYTCIVENTAGKDTCEATLVIESLEKKSEKAPFEFVALQDKTKTKSEK 3011
QY 884 T-----PETTTAAPKMTKETATTTEKT--TESKITATTQ-----VTSTTQDTTFF 928
Db 3012 VLECKVIGPKFVSWLHNVNREKPNSEKTIQESITVESVEGVERVTITSE----- 3066

```

```
QY 929 KITT-----KTTLAPKVTTKT-----ITTTIMMKP-----958
Db 3067 -----LSHQKYTCIAENTGTSKTEAPLTVQGEAPVFTKELONKELSIGEKLVLSGVK 3122
QY 959 -----ESTAKPKDR-----ATNS 971
Db 3123 SPOPHVDFYSFSETTKVETKITSSRIAIEHDQTNTHRMVISOITKEDIVSVKATINS 3182
QY 972 KATTPKPKPT---KAP---KKPTSTK-----KP-----994
Db 3183 IGTATSTSKITTKVEAPVFGQLKKTSVKEKEIKMEVKVGGAPDVEWFKDKPVSEGD 3242
QY 995 -----KTMPR-----VRKPKTT-----PT-PRKM 1012
Db 3243 NHEMKKNPETGVFTLVVQKAATTDAGKYTAKASNPAGTAESSABAEVTSLEKPTFVREL 3302
QY 1013 TSTMPELNPTSRIAEAMLOTTTR-----PNOTENSKLV-----1045
Db 3303 VTEVKINET-----ATLSVTVKGVPPSPVSEWLKOGQPVQTDSSHVIAKVEGSGSYSITI 3357
QY 1046 -----EVNPKSEDAGAGETPHMLLR---PHVFMPEVTP-----1077
Db 3358 KDARLEDSGKYACRATNP---AGEAKTEANFAVVKNLVPPPEFVEKLSPLEVKEKESTTL 3413
QY 1078 -----DMD-----YLPVPVNO 1088
Db 3414 SVKVVGTPSPVSEWFKDDTPISIDNVHVIOQTAVGSFSLTINDARQDVGIVSCRARNE 3473
QY 1089 -----GII-----INPM-----LSDETNIC 1103
Db 3474 AGEALTTANFGLIIRDSIPPEFTQKLRPLEVREBQETLDLKVTVIGTPVNVFWFKDD---3529
QY 1104 NGKEVD-----GLTTLRNGTLVAFRGH-----1125
Db 3530 --XFINIDNSHIPAKDESGHHTL--TIKQARGEDVGVTCKATNEAGEAKTTANMAVQ 3584
QY 1126 -----YFWMLSPF---SPSPARRITVWGPISPI-----1152
Db 3585 EEIEAPLFGQLKPYEVEQGKPAELVVRVEGKPEPEVKWFGDGVPIAIDNOHVIEKKGEN 3644
QY 1153 -----TN-DIK-----DAGY 1185
Db 3645 GSHTLVIKDTNNADFGKVTQATNKAGKDETVGBLKIPKPSFKETAEEVVKPLFIEPLKE 3704
QY 1153 -----DTVFRPCNCEGKTF---FFKDSQYWRP-----1176
Db 3705 TFAVEGDTVVLECKVKNKESHPIKFFKNDQPVIEGQHMQLVLEDDGNIKLTIONAKKEDV 3764
QY 1177 -----TN-DIK-----DAGY 1185
Db 3765 GAYCEAVNVAGKANTNADLKIQFAAKVEHVHVDSEGOLEBIGQFETVGDFTASSKTDTGR 3824
QY 1186 PKPIF-----KYKNWPE---SVYFFKRGGSIO-----KGFGGLT 1197
Db 3825 GAPEFVELLRSCVTVEKQQAILCKVKGEPRPKIKWTKEGVEMSARVRAEHKDDGILT 3884
QY 1198 -----GOIVAAIISTA-----1207
Db 3885 LTFDNTVTOADAGEYRCEAENEYGSATTEGPIIVTLEGAPKIDGEAPDFLPVFPVAVTVG 3944
QY 1208 -----KYKNWPE---SVYFFKRGGSIO-----OY---1228
Db 3945 ETAVLEGKISGKPKSPSVKWKYNGEBELKPSDRVKIENLDGQTQRLTVTNAKLDDMDVEYCE 4004
QY 1229 -----IYKQFQVQKPCGRREALNPVYVGEHTQVRRRPFERA-----1264
Db 4005 ASNEEGDVMSDVLTVKEPAQVAPG-----FPKELSAIQVKETETAKFECKVSGTKP 4056
QY 1265 -----IGPSQTH---TIRQYSPARLAYQDKGVILH 1291
Db 4057 DVKWFKDGTPLEKDKRVHFESTDDGTQRLVIEDSKTDQGNRYIEVS-----NDAGVAN 4110
```

```
QY 1292 NEVKVSI-----LWRGLPNV-----VTSA-----1310
Db 4111 SKYPLIVVSETLKIKKGLTDVNTQGTKILLSVEVEGPKTKVWKYKGTETVTSQTTKI 4170
QY 1311 -----ISLPMIRK-----1318
Db 4171 VQVTESEYKLEIESAEMSDTGAYRVVLSTDSFVSSESSATVTVTKAAEKISLPSFKKGLAD 4230
QY 1319 -----PD-----1320
Db 4231 QSVPKGTPLVLEVEIEGKPKDVKVYKNGDEIKDKVEDLNGKYRLTIPDFOEKDGEYS 4290
QY 1321 -----GYDYIAFSK-----1329
Db 4291 VTAANEAGETESKAKVNVNSAKPELVGLVPTTVKQGETATFNKVGKVPKVKYKNGKE 4350
QY 1330 -----DOYVNIIDVP-----SRTARAI-----1345
Db 4351 IPDAKTKDNGDGSYSLEIPNAQVEDAADYKVVNSNDAGDADSSAALTVKLADGDKVKVP 4410
QY 1346 -----TTRSGQTLS---KV-----WY 1358
Db 4411 EIVSGLIPTTVKQGETATFNKVGKVPKVKQWKY 4443

RESULT 10
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR0199
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S48478; A26877; B26877; S27281; J06123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364;
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YAZ>
A;Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccha
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudoh
A;Reference number: J06123; MUID:96323237; PMID:8710886
A;Accession: J06123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:gl304386; PIDN:AAC49609.1; PID:gl304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001456
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
```


C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match	13.3%;	Score 971;	DB 1;	Length 1367;
Best Local Similarity	22.3%;	Pred. No. 7.9e-13;		
Matches 377;	Conservative 128;	Mismatches 487;	Indels 700;	Gaps 76;

QY	13	LLSVFVITQVSSQDLS	SCAGRC	GGYGRDATC	-----NCDYN-----	49
Db	6	LLAYLVLSLLFN	SALGFPTAL	VPRGSEGTSC	NGIVNGCPNLDFNWHMQDQNI	65
QY	50	-----	-----	-----	-----	49
Db	66	TSVSWQDNTYQIT	IIHWKGENIDL	KYLSLKIIGVTG	PKGTQVLYGYNENTY	125
QY	50	-----	COHYMECCPD	KRVCTAEL	SCKRCPE	76
Db	126	FTATFEVATQD	VNSCQVM	---PNFOIQ---	PEYLGSAQYASSWQ	171
QY	77	SFERGRC	-----	DCDAQC	-----	90
Db	172	SFDLSTGCNNY	DQGHQSOTD	PGFYWNIDCNN	CGTKSSTTSS	231
QY	91	-----	KKYDKC	CPDYESFCAE	VHNPTSPSSKKAP	130
Db	232	TTTSSSTSE	STSSSTTAP	-----	ATPTTSC	278
QY	131	STTKRS	PKPN-----	KKTKK	VESEBITEVKN	183
Db	279	SCTK	KEKTPPHD	TPCTKKKT	-----	315
QY	184	DNGDFK	VTTPDSTTQ	HNVSTSPK	ITAKPINPRPS	240
Db	316	-----	TPSSST	-----	SSAPVPT	352
QY	241	---VETKE	TTTNKQTS	---TDGKE	TTSAKETOSIEK	295
Db	353	SAPV	PTPSSST	TESSAPV	TSSTESSAPV	403
QY	296	TKGPA	-----	LTTPKE	PPTTPKEP	342
Db	404	ESS	SAPV	TSSTESSAPV	TSSTESSAPV	459
QY	343	TKS	APTTPKEP	APTTPKEP	APTTPKEP	402
Db	460	TP	SSSTESSAPV	TS	-----	511
QY	403	-----	TTTPKEP	APTTPKEP	TP	452
Db	512	VPT	PSSSTESSAPA	---PT	SSSTESSAPV	560
QY	453	KEP	APTP-----	KEP	APTTPKEP	503
Db	561	TT	ESSSTPV	TSSTESSAPV	TPS	617
QY	504	TP	KEP	SPPTTPKEP	APTTPKEP	562
Db	618	TT	ESSAPV	TS	-----	669
QY	563	KEP	APTTPKET	APTTPK	KLAPTTPK	619
Db	670	TP	SSSTESSAPV	---	SSSTESSAPV	717
QY	620	EP	APT	TPKAAANT	TPKEP	672
Db	718	SS	APVPTP	SSSTESSAPV	TPS	773
QY	673	TL	KEP	APTTPK	PAKELAP	725
Db	774	SS	SAPVPT	SSSTESSAPV	TP	830
QY	726	-----	TPKEP	APTTPK	GAT	768

Query Match	12.7%; Score 931.3; DB 2; Length 4135;	
Best local Similarity	12.3%; Pred. No. 6e-11;	
Matches 462; Conservative 197; Mismatches 496; Indels 2597; Gaps 125;		
QY 23	SSODLSS--CAGRC-GEYSRDATCNDY-----NCQHYMEC-----	56
DB 518	AGEDCGSRFCGDCRGRCGDGVCSDVGEYEGDCGRKSCPRGCGQCGLEGRVCDD	577
QY 57	-----CPD-----	59
DB 578	GYEGEDCGVRRCPDCNQRGVCQGVCTCWEGFAGEDGLRVCPSNCHRRRCENGRCVC	637
QY 60	-----PKVCTAELSCGR-----	73
DB 638	DSGYTGSCATRTCPAD--CRGRGRCVQGVCHVGYGEGDCGQEBPPASACPGCGCPRE	695
QY 74	-----CPESFERGREG--DCDAOCKYDKC-----CPD-----	99
DB 696	LCSAGOCVCVEGF-RGPDCAIQTCPCDCRGCEGRCSCQDGVYAGEDCGEEVPAIEGM	754
QY 100	-----YE-----SFCABVNPIS-----	112
DB 755	RMHLEBETVTRTEWTRAPGNVDAYEIQFIPTTEGASPPFTARVPSSASAYDQRLAPGQE	814
QY 113	-----PSSK-----	117
DB 815	YQVTVRALRGINWGPASKTITIMDGPDLRVVAVPTTLELNLWLRQAEDRFVSYV	874
QY 118	-----	117
DB 875	SAGNRVRLEVPSEADGTLTGLMPGVEYVVTVAERGRAVSYPASIRANTGSSLSGLLG	934
QY 118	--KAPPSGASQTKSTTKSPKPNKKTKKVIIESEBIIEVKONKNRYKK-----	167
DB 935	ATDEPPPSG--PSTTQGAQAPVLOQR-----POELAEIRVLGDKDKTGLRLVAWTAQ	983
QY 168	-----KPTPKPP-----	174
DB 984	PDFTFTHQLRLRVPEGGAHEHLLPGDVRQALVSPPPESPEYELSLRGIPSGPSAPLI	1043
QY 175	----VVDGAGS-----GLDNGDF-----	188
DB 1044	YQIMDKDGEKPGKPLAPPLRGKLTVDVTSLSLLHWTVEGEFDSFVIOYKDRDRPQV	1103
QY 189	KVTTPTDTSITQHN-----KVSISPKITTAQPI--NPRSLPN-----	224
DB 1104	VPVEGPQRSALISNLDVGRKYKFLYGLVGRKRGHGLVAEAKILISQTDPSVTPPLGNL	1163
QY 225	-----	224
DB 1164	WTDPTDLSLHLSWTVPEGQFDSFMVQYRDRAGRQVVPVVEGPDERSVLIISPLDPKHYP	1223
QY 225	-----SPTSKESTSLVNKETTETKE-----TTTTNKQTS-----	254
DB 1224	TLFGIANKRHGPLTADGTTAPEKKEPRHPEPPERPLLGLTAVAGATADSLRLSWTVAQ	1283
QY 255	-----TDGKEKTSASETSIEKTSADKLAPTSK-----VLA 286	
DB 1284	GSFDSFVVQYKDAQRPQAVPVTGDENEVAIPSLPEPRKYKMNLYGLHGRORVGPVSVVA	1343
QY 287	KPTPKAETTTKGPALTTPKE--PTTTPKEPASTTPKE-----TPTTKSAPTTP	335
DB 1344	TTAQ--EVLDETPSATMEETPSFTEPSTKAPESPEKELLGELMVTGSSPDSLSLWIP	1402
QY 336	K-----EP-----	338
DB 1403	QGHFDSFTVQYRDGQPRVNRVPGDEDEGVTISGLEPDHKYKMNLYGPHDRQVRGPMMSVI	1462
QY 339	APTTKSAPTTPKE--PAPTTKE--PAPTTKEPAPTTPKE--APTTKSAP-----	386
DB 1463	GVTABEETPSFTEMEETPSFTEVEETSPMEPSTEAPPEAPPEPLLGLTWTGSSPDSLS	1522
QY 387	---TTPK-----EP-----	392
DB 1523	LSWTVFQGHFDSFTVQYKGRDGPQVVRVGRGEETEVTIGGLEPGRKYKMNLYGLHGGRSG	1582
QY 393	-----APTTPKAPAPTTPKEPAPTTPKE-----TPTTPKE-----	423
DB 1583	PVSTVGVTAPQABETPPATEP-----PKEPRILGELTVDVTNSVGLWMTVSEGOFDS	1635
QY 424	-----PAPTTKEPAPTTPKE-----APTAPK 445	
DB 1636	FMVQYKDRDQCQSHVVPVVAADQREATVSGLEPERKYMNVYGLHGGQVGLSVVALLAPV	1695
QY 446	KPAPTTP--PKEP-----APTTPKE-----PAPTTP 467	
DB 1696	PPDPVTPPVPVPRILGELTVDVTIPDSVGLSWTVAEGEFDLSFLVOYKDRDGPQVVPVATD	1755
QY 468	TKE-----PSTTP-----KEPA 480	
DB 1756	QREVTIGLEBFSRKYKFLFGIQDKRKRSTVSVEAKTVGRGDASPGAPPLGELMWTDPT	1815
QY 481	PTTKSAPTTP-----	490
DB 1816	PDSLRLSWTVPEGHFDSFVVQVKDRDGPVVSVEGHERSVTISPLDSGRKRYRFLVYGLLG	1875
QY 491	TKEPAPTTPKSAPTTPKEPSPTTPKEPAP-----TTPK-----	523
DB 1876	KRRHGPLTTGGTTTTRAVDEAGTKRPSKPRLGELOQVGTNSVGLSWTVPEGHFDSF	1935
QY 524	-----BPAP 527	
DB 1936	VIQYRDGQGVVPVVEGSRREVSVGLDPARRYKLLYGLSRDKRVGPISAIATVPAP	1995
QY 528	TTPKAPAPTTPKEPA-----PPTPKEPAP-----TTTK-----	556
DB 1996	REEIKAEPAIFSPASPPLLEGEVTLLEAAHSLRLSWTATEGEFDSFEVQYTDENGLOQE	2055
QY 557	-----PAPTAPK-----EPAPT-----	568
DB 2056	VNVGQDQHDITISLSDHRVLYSLYGFHDQVRGPAHIEAMTAPREDDPESLSLSTQ	2115
QY 569	TPKETAP-----TTPKLTTP--TTP-----	587
DB 2116	TPSTAVPEPHIKPRLGELAVTDTTPDSLSLSWTVPEGOFHFLIOYKNGDQPKVVRVP	2175
QY 588	-----KLAP-----TTPKAPAPTTPPELAPTTP--P 613	
DB 2176	DEDEVITISGLEPDHKYKMNLYGFHNRQRMGVSIVIGVTABEETP-----SPTMEETP	2229
QY 614	TPTTPEE--PAPTTPKAAAPNTPKP-----	637
DB 2230	SPTMEETPSFTEPTEVPPEPPEPPELLELTVTGSSPDSLSWTVPGHFDSEFTVQYKN	2289
QY 638	-----APTTPKE--PAPT 648	
DB 2290	GDGQPKVVRVPGHDKGVTISGLEPDHKYKMNLYGFHNRQVRGVSIVIGVTABEETPSPT	2349
QY 649	TPKE--PAPTTPKETAPTTPK-----GTAPTTL-----	674
DB 2350	EVEETPSFTEPSTEAPPEPPEPPEPVLGELMVTGSSPDSLSWTVPGHFDSEFTVQYKRDG	2409
QY 675	-----KEPAPTTPKPP 685	
DB 2410	POVVRVGEETEVTVEGLEPGRKYKMNLYGLHGRVRGVPSTVAMTAREEPPASPPLKP	2469
QY 686	APKELAPT-----TTPK-----	697
DB 2470	QLGELTVDATPDSLSLSWTVPEGQFDFHFLVOYKNGDQPKAVRVPGEDEGVTISGLEPD	2529
QY 698	-----PTS-----TTSK-----PAPTTPKGTAPTTPKE-----	722
DB 2530	HKYKMNLYGFHNRQVRGVSIVGLTVSEKQDQEMTAPDTDLPTAAPEPEIKPRLGELVVD	2589
QY 723	-----APTTPKEPAPTTPKGTAP-----	740

Db 2590 ATPDLSLSWTVPEGQDFHFLIQKNGDQPKAVRVPQDGEDEVITISGLEPDHKYKMNLYG 2649
 QY 741 ---TTLKE--PAPTPKK--PAKELAPTTTKGPTSTTSKPAPTPKPE 782
 Db 2650 FHDQRGMPVSVIGVTTABEETPGTEMEQTPSPTEVEET--PGPTE-VEETPSPTEPST 2706
 QY 783 TAPTPKPE----- 790
 Db 2707 EAPRPEEPLLGELTVTGSSPDSLSWTVPGHFDSTIOHKGRDGPQVVRVCGEETEV 2766
 QY 791 -----PAPTPKKPAPTTPKP- 807
 Db 2767 TIGGLEPGRKYKMNLYGLHSGQRVGPVSTGVTDQEVVEETPSPTEPSTEAPPEEPL 2826
 QY 808 ----- 807
 Db 2827 LGDLTVTGSSPDSLSWTVPGHFDSTIOYKGRDRPQVVRVCGEETEVITIGLEPRHK 2886
 QY 808 -----PPTTSEVSTP---TTTKBPTTIHKSPDESTPELSAEP----- 841
 Db 2887 YKMNLYGLHSGRRVGPVSTGVGTAPDYDAMTTQTPST-----SVPEPTPKPRLGEVTV 2939
 QY 842 ---TPKAL-----ENSPK-----EP----- 853
 Db 2940 TDATPDSLSWTVPEGQDFHFLVQYKNGDQPKAVRVPQDGEDEVITISGLEPDHKYKMNLY 2999
 QY 854 -----GVPTT--KTPAAKPEMTTAKDKTTERDLRTPTTTTAAPKMTK 896
 Db 3000 YGFHDHVRGPVSVIGVTTABEETPSPTEPSTEAPPEEPLLGELTVTGSSPDSLSWTVPGHFDSTVQYKGRDGPV 3039
 QY 897 ETATTTKTTES-----KITAT----- 913
 Db 3040 ETPSPTEPSTEAPPEEPLLGELTVTGSSPDSLSWTVPGHFDSTVQYKGRDGPV 3099
 QY 914 ---TTQVT----- 918
 Db 3100 VRVGGEETVTVGLEPGRKYKMNLYGLHSGRRVGPASTVGTASLATERPLAPRLGELA 3159
 QY 919 -STTTQDTT---PF-----KITLKTITLAP----- 940
 Db 3160 VAVVTSDBLSWTVPEGQDFDSFLVQYKVDQVQVQVAGDVRETVSSLPAGRKYKFL 3219
 QY 941 -----KVTTTKTITTEIMNKPETAKP----- 964
 Db 3220 LFGLRDEKRGHPVSADAKTLPTD-----KPAERLGLTVDVTPSGVGLSWTVPE 3269
 QY 965 -----KDR-----ATN 970
 Db 3270 GFDFSMVQYKDRDQGPVVPVVAADQREVTVPGLEPNRKYKFLYGLVGRKRLGPIAEG 3329
 QY 971 SKATTPKPKP----- 981
 Db 3330 STAPLEKERPPRLGELTVTDETPNSLSWTVVAQGRFDSFVVQYRGDGPQPMVPA 3389
 QY 982 -----TKAPKPKP----- 988
 Db 3390 DOREFTVEGLEPGRKYKFLYGLLGGQRLGPASVLGWTAEEDTPAPWHAATEAPKPEG 3449
 QY 989 ----- 988
 Db 3450 FRLGVLAVRDVSPLSLRSLWSVVQGFDSFVVQYQDTDGPQALLVGGDQNKVLVSGLEP 3509
 QY 989 ----- 988
 Db 3510 STSYEFYGLHGEKRLGPVSAEGTGTGPVAGOTPGEPGRLSLWTDVTTSSLRNWE 3569
 QY 989 -----TSTKPKMTPRVRKPKTTPTRKMTSTMPELNP----- 1021
 Db 3570 APPEAFDPLRFGVPSFSTLEQLRLLQRELTVPGTRR-SAVLRDLHFGILYTLTYLG 3628
 QY 1022 -----TSRIAEAML-----QTTTRPNQTPNSKLVE---VNPXSEDA 1054
 Db 3629 LRGPBKADSIQGTARTLSPVLESFRLQFSEIRTSARVSWTPTSRVDGFKVSYQLADG 3688

QY 1055 G-----CAGCETHMLLRPHVFMPEVTPDMDYLRPRVNOQIINPM 1095
 Db 3689 GEPOSQVQDGRQKLEGLIPGAQYEVTVSVRG---FEESEPLTGFITVDP-----GPT 3740
 QY 1096 ---LSDETNICNGK-----PVDGLTTLN- 1116
 Db 3741 HLRALNUTDSALUHWPPQTPVDYDVKVTAAGAPSLQASAPGSAVDYPLQGLVTHNTY 3800
 QY 1117 -GTLVAFRGRHYF-----WMLSPFSPSP-----S 1137
 Db 3801 TAILRGLRGNFTSPASITFTTGLEAPQDLEAKEVTPRTALLTWTAEVSTPGYLLSNT 3860
 QY 1138 PARRITEV-----WG-----IPSPID 1153
 Db 3861 FGGQTQBIILLPGGVTSQHLRGLFPSTPYSTWLRAMWGDSTPPVSTSTFTTGLRLPFRPD 3920
 QY 1154 ---TVFTRCNCE-----GKTFEFKDSQYW 1174
 Db 3921 CGEEMQNGVSTRTTIFLNGNRERPLNVFCMDTGGWLVFORMDGKTDFWRD---W 3977
 QY 1175 RFTNDIKDAGYKPKIFKFGSLTGQIYAALSTAKYKNW--PESVYFFRKGSIQOYIYKQ 1232
 Db 3978 ---EDYAH---GFGNISGEF-----WLGNEALHSLTKAGD----- 4006
 QY 1233 EPVOKCPRRPAALNYPVYGEWTVRRRRFRRAIGPSOTHIRIQYSPARLAYQDKGVLHN 1292
 Db 4007 ---YSLRVD-----LRAGEAFAVFAQ 4023
 QY 1293 EVKVSILWRLGPNVVTSAISLPNIRKPDGYD-VYAFSKDQYVNDIVPSRTARA---ITTR 1348
 Db 4024 ---YDSFOVDSADEYRULHLESGYHGTAGDSMSYH 4054
 QY 1349 SGOTLS-----KVWY-NC 1360
 Db 4055 SGVSFSARDRPNMLLISCAVSRYGAWMYRNC 4086

RESULT 12

S49915

extensin-like protein - maize

C/Species: Zea mays (maize)

C/Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004

R/Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A/Description: Pex genes: pollen-specific genes with extensin-like domains.

A/Reference number: S49915

A/Accession: S49915

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1188 <RUB>

A/Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g6001

Query Match 12.5%; Score 917.8; DB 2; Length 1188;

Best Local Similarity 18.9%; Pred. No. 7.9e-12;

Matches 299; Conservative 81; Mismatches 397; Indels 802; Gaps 57;

QY 10 LLLL---LSVFVITQVSSQDLSSCA-----GRGCE----- 36
 Db 17 LLLLAACLACSQAVTSAEASVIAHRQLLAMKEAGGEGADLPADFEFDDRVGAANFPN 76
 QY 37 -----GYSRDATCN-----CDYNQOHYMECCPDFFKRVCTAEL 68
 Db 77 PRLRRAYIALQAWHRAFYSDPKGY---TANWVGEDVCKYNGVICTEALDDPKITVVAGI 132
 QY 69 SCKGR----- 73
 Db 133 DLNGADIAGYLPPELGLLTDLAFFHTNRFQGIIPKMSRSLSLLEHFDVSNRNFVGVFP 192
 QY 74 --CFES-----FERGRECCDCA----- 88
 Db 193 YVCLEWVSLKYLDLRFNDFEGELPPALFDK---DLDAIFVNTNRFVGPINLGNSTAS 248

Query Match			
Best Local Similarity 11.34; Score 916.8; DB 2; Length 5762;			
Matches 535; Conservative 197; Mismatches 474; Indels 3512; Gaps 155;			
Qy	2	AWKTLPIYLL---LLSVFVIQVSS-----	QDLSS 29
Db	753	AHNSVTDMFLSPVDLQIFRSTQSKYKTIKIHEDLALITIPESLSDGSLTFQEDLLQ 812	
Qy	30	CAGRCGSGYSDATCNDYCNQHWECPPDKRVCTAELS-----CKGRCF-----ES 77	
Db	813	PIDSTGGQ-----BFSHIKTSLKPPPYVSNTKSSAFOETMSET 851	
Qy	78	FEREGCDCAQCKYDKCCPDY-----ESFCAEVH-NPTSPSSKKAPPS 123	
Db	852	TYSNQVDLSSTHLKTSLELDPDYMGLEPSLYQOITQSSPKSLHENPKSP-----902	
Qy	124	GASOTIKSTTKRSKPKPNKK-----	144
Db	903	---VLKPSAAQALKPRKEKLSPTNNMIPHPKSLKNNMVTHIPAHKMTVPRQIQEDQG 958	
Qy	145	-----TKVIESEBEITEV-----DNKK 162	
Db	959	EYTISSNGSQPLDLEVLTSGLIPEVKHILPKRTVNPQYSQVKISHSQHVETQHENSE 1018	
Qy	163	NRTKK-----KPTPKPPVVDEAGSLDNGDFKVTTPDTS-----197	
Db	1019	TTIVQPLDLEFALNQLTPKE-----NFAQTQDITTMIGPPEKVEIAQPE 1065	
Qy	198	-----INPRPS-----TOHKK 202	
Db	1066	HHEGTIPIQDAEYSTLPTVSFQPLDQLBITTSEAIRBPHPTVPQQIIVHPPHPL 1125	
Qy	203	VSTSPKITTAKP-----	220
Db	1126	VIHSEQVYTOHPNTEALIQPLDLELITPQTAEGELPQLDQSTTQIIEPPTVVVGPV 1185	
Qy	221	-----LPPN---SDTSKETSITVNKETTET-----KETTTN---250	
Db	1186	PIYEEVTVQITSQDAEYRPSPTVSFQSLDLELITISPEATREYHPSLLQOITIVNPPH 1245	
Qy	251	-----KQSTSD-----GKEKTTSAKETQSIKTSAMDLAPTSKVLA 286	
Db	1246	PLVIHSEQVHTQHPNLTEATVQPLDPLTITPQTTGEGELPQLQDS-----TSQIII 1298	
Qy	287	KPTPK-----AETTKGPALTTP-----KEP-TPPTPKPEAST 318	
Db	1299	EPPIKVVALVPVYQEVSQDAEYTTSTVSFQPLDQELTITSAIRBPHPTVPQOITIV 1358	
Qy	319	TP-----KBPTPTTIKSAP-----TPPKPA-----PTTKSAPTTPKEPAPT 356	
Db	1359	HPTKHPLVIIHSEQTOHPNTEVTVQPLDLELITPQTAEGELPQLDQSTTQIIEP-PT 1417	
Qy	357	TTKEPAP-----TPPKE-----PAPTT-----373	
Db	1418	VVVGPPVPIYEEVTVQTSQDAEYRPSPTVSFQSLDGLTITPEPTEHFITQKTIVFPPM 1477	
Qy	374	-----TKEPAP-----TTTKSAPTTPKEP-----392	
Db	1478	YTDVTLPOQVSVQHLKPTGIVQPLDLELITPQPTPEGELSQTVQESTTONKEPKHEVV 1537	
Qy	393	APT-----TPPK-----399	
Db	1538	APFVVQAVTVPTPSQVQAEYKSLQPLDLELITVTSPTKEAVHSTISKNSLAINPQYVH 1597	
Qy	400	-----PAPTT-----PKE-----407	
Db	1598	IOHPNPAEATVQPLDLELITSSLOPTABGELLYSMQETVTQISEPPKQVVPVPEYQEV 1657	
Qy	408	---PAP-----TTPKETPTTPKE-----423	
Db	1658	AVPAPVQDAKYPLSSIVSLNSLDQELTSLSELLGEAHLQITTPDETMTWLPKPKDROGIYD 1717	

Qy	424	-----	423
Db	1718	HDHKKHLNLTETVNTQPFHLEHTVQHQPTIEERSQSICKKTTQITEPGKKVVPLAQESEE 1777	
Qy	424	---PAPTTKEPAPTT-----	436
Db	1778	VTIPMPILKETAPFTPHSMALQSLDEKLTIHSHSPGWTQQHANLKESKGHTTGKILLDYA 1837	
Qy	437	-----	K 437
Db	1838	EPNMEIELKHGLLTKTTEATTESENTNQMTSLKQVLTALFTQNKKSMALPALVESQDES 1897	
Qy	438	EPAPTAPKPA-----PTPK-----EPAPT-----458	
Db	1898	QPPNMSLQPLDQELTSLSQPHGMVPHIPNTPEKIYLYHAEP-PTGPFVBPPELFLKTT 1956	
Qy	459	---TPKE-----PAP-----TTTK 469	
Db	1957	KSRPVQGTATQMAASPKEYMYSRAPENKEAVLSGEGEDQDESPPNMSLQSLDQELTSS 2016	
Qy	470	EPSPTTPKEP-----APT-----TTKSAP-----488	
Db	2017	QPHGWIPHPNTHGKIYLYHAEPPTGPFVBPDLFLKTTKSPVEVWTLRTDKSRKEMV 2076	
Qy	489	-----TTTKEP-----494	
Db	2077	SQSPKYEEAVLPVHGQEESRPPNMSLOSLEQLTSLSQPHGMVPHPNTHGKIYLYH 2136	
Qy	495	---APT-----TTKSAP-----503	
Db	2137	AEPPTGPFVBPDLFLRTTKSKPVQGTATRMVKSPEEMVSLDPENKEAVPPAQGEKG 2196	
Qy	504	-----TTT-----TT 505	
Db	2197	SPSSPNMSLQSLDHELFMSQPHGWIHPBPCTDKIYLYHAEPPTGPFVBPDLFLRTT 2256	
Qy	506	PKESPSTTKKEPAPTTPKSPAPTT-----KKAP-----535	
Db	2257	KSKPVQGTITTEMA-KSPKEMVSQTPYEKAVLSGEGEDQDESPPNTSLKSLDQEVAMS 2315	
Qy	536	TTKEPAPTPK-----EPAP-----TTTK-KPAPTAP-----KE 564	
Db	2316	SQSHSGVPHPKTPGKIYLYHSIEPPGPFVKPTDILLVKTITKSPAETWPRIDKLLKE 2375	
Qy	565	PAPTTPK-ETA-----PTTKKL--581	
Db	2376	MVPHSPEYEEAVFPAHGEQDESPPNMPLQPLDQELTSLSQPHGMVPHPNTPGKIYL 2435	
Qy	582	---TPTTPEKLAP-----TTPEKAPTTPELAPTTEPTPTTP-----618	
Db	2436	HYAEPPTGPFVBPDLFLRTTKSKPVQSGPKEIA-KSPKEMVSQTPYEKAVLSGEGED 2494	
Qy	619	---EEPAP-----TTPKAAAPNTPK-----EPAP-----639	
Db	2495	QDESPPNMSLQSLDQEVMTSSQPHSGVPHPKTPGKIYLYHSIEPPGPFVKPTDILLV 2554	
Qy	640	---TTPEKAPTT-----KEPAPTTPK-ETA-----662	
Db	2555	KTTTKSPAETWPRIDKLLKEMVPHSPEYEEAVFPAHGEQDESPPNMPLQPLDQEL 2614	
Qy	663	-----PTTP-----KGTAPTTL--674	
Db	2615	TLSSQPHGMVPHPNTPGKIYLYHAEPPTGPFVBPDLFLRTPKSKPVQGT-PTQMAKS 2673	
Qy	675	-----KEPAP-----679	
Db	2674	PSEMVSLSPKNETVFPAGCKQDESPPNLSLQSLDQBITMSSQPHGWIHPNTHGK 2733	
Qy	680	-----TTPKP-----APKEL-----690	
Db	2734	IYLYHAEPPTGPFVBPDLFLKTTKSKFMQGSPRQIDKSPKEMFTQSPYEESLLPAHA 2793	
Qy	691	-----AP-----692	

```

Db      2794  EGQESRAPPHFSLQPLDQELSUSHPHGWIPHHNPNTDKIYLHYABPTGPFVEPPDLF 2853
QY      693  ---TTTKEP---TSTTSK-----PAP-----708
Db      2854  FLKTTKSKPVQGTATKTDSPEDRVSTQTPYKEAVLSGPGDQDESPPNMSLSQLDQE 2913
QY      709  ---TTPKGTAPTPKEP-----APT---T 726
Db      2914  LAISSQPHGWIPHSNAPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPLQGTQMAKS 2973
QY      727  PKEPATPK-----736
Db      2974  PKEMVSTPYKADIASAPGENDESPPNMSLSLHLDQELSLSQPHGWIPHPNTHGK 3033
QY      737  ---GTAP---TTLKEP-----746
Db      3034  IYLHYABPTGPFVEPPDLFLLKTTKSKPVQGSQIDKSPKEVFSQSPSESVLPAQA 3093
QY      747  ---AP-----748
Db      3094  EGQESRAPPHMSLQLLDQKLTLSSHPHGWIPHPNTPGKIYLHYABPTGPFVEPPDLF 3153
QY      749  ---TTPKPB---APKEL---APTITKG-----766
Db      3154  FLKTTKSKPVQGSQIDKSPKEVFSQSPSESVLPAQAEGQESRAPPHMSLQPLDQD 3213
QY      767  ---PTS---TTSKPAPTTP-----780
Db      3214  LTLSSHPHGWITHPNTPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPAQTPTQMAKS 3273
QY      781  ---KETA-----784
Db      3274  PEMVSLSPENKETVFPQAQGOESISPPHMSLQPLDQDLTPSSHPHGWIPHPNTHGK 3333
QY      785  ---PT---TPKE---PA-----792
Db      3334  IYLHYABPTGPFVEPPDLFLLKTTKSKPVQSPRQIDKSPKEVFTQSPYEESVLPAQA 3393
QY      793  ---792
Db      3394  EGQESRAPPHMSLQPLDQDLTLSSHPHGWITHPNTHGKIYLHYABPTGPFVEPPDLF 3453
QY      793  ---PTTPKPA-----800
Db      3454  FLKTTKSKPVHGSQPRQIDKSHKEMFTQSPYEESVLPAQAQGOESRAPPHMSLQPLDQD 3513
QY      801  ---PTTPEP---PPTTSEVSTP---TTTK---EPTTIHKS 829
Db      3514  LTLSSHPHGWIPHPNTPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPVQGSQIDKSK 3573
QY      830  PDE---STPLSAEP-----841
Db      3574  PKEMFTQSPYEESVLPAQAQGOESRAPPHMSLQPLDQDLTLSSHPHGWIPHPNTPDK 3633
QY      842  ---TPKALENSPKE-----852
Db      3634  IYLHYABPTGPFVEPPDLFLLKTTKSKPVHGSQPRQIDKSPKEVFTQSPYEESVLPAQA 3693
QY      853  ---PGVP-----856
Db      3694  EQQESRAPPHMSLQPLDQDLTLSSHPHGWITHPNTPDKIYLHYABPTGPFVEPPDLF 3753
QY      857  ---TTTKT---PA-----862
Db      3754  FLKTTKSKPVHGSQPRQIDKSPKEVFTQSPYEESVLPAQAQGOESRAPPHMSLQPLDQD 3813
QY      863  ---ATKP-EMTITAKDKT-----876
Db      3814  LTLSSHPHGWITHPNTHGKIYLHYABPTGPFVEPPDLFLLKTTKSKPVQWTPTQIDKS 3873
QY      877  TERDLRTPE-----886

```

```

Db      3874  REEMVQSPENEEADIPRHGDQDELNRNPEIISLQPLDQELTLSSQPHGWVPHPPSTPDK 3933
QY      887  ---TTTAAP-----KMTKETATTE-----903
Db      3934  IYLHYABPTGPFVEPPDLFLLKTTKSKPVQGTPTQLAKSPKEMVFTQTPYKEAVLSAPG 3993
QY      904  ---903
Db      3994  EDOETSPSPNMSLQSLDQELTMSSQPHGWLPHPPNTHGKIYLHYADPTGPFVEPPDLF 4053
QY      904  ---KTTESK-ITATTTQVSTTTQ-----923
Db      4054  FLKTTKSKPVQGTPTQMDKSPKQGLLPAHGEAQDESPPNMSLQPLDQELSLSQPHGW 4113
QY      924  ---DTPPKITT---LKT-----936
Db      4114  ITHPPNTPGKIYLHYABPTGPFVEPPDLFLLKTTKSKPLAGTPTQMAKSPKEMFSOTLE 4173
QY      937  ---TLAP-----940
Db      4174  HKEAVLSAPGEDQDESPPSSNMSLSQLDQDITLSPQOYGRIPHPVTSQKIYLHYVEHPT 4233
QY      941  ---KVTITK-TITTEIMNKPE-----960
Db      4234  VPLFQHPNMFSTKTTTTSKPLQLTQTOMAKSPKEIVSLRPEYKEAILPAQVESQESVLP 4293
QY      961  ---TAKPD-----966
Db      4294  NMSLQSLDQELTLSSQPHGWIPHPNTPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPV 4353
QY      967  ---966
Db      4354  QWTPTQIDKSLKXVAGSPYEBAVPAHGEQDESRYAENISLSQLDHDITLSSQPHGW 4413
QY      967  ---RATNSKAT--TPKP-----978
Db      4414  IPHPNTPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPAQTPTPTMAKSPPEMVSLSK 4473
QY      979  ---OKPT-----982
Db      4474  NKEAVLPAQKDBEESISPPNMSLQPLDQELTLSSQPHGWVHPPNTHGKIYLHYABPT 4533
QY      983  ---KAPK-----987
Db      4534  GPFVEPPDLFLLKTTKSKPVQGTITQMVKSPKQGLLPAHGEAQDESPPNMSLQPVQDE 4593
QY      988  ---PTST---KPKXTMPVRKP---KT---TPT-----1008
Db      4594  LSLSSQPHGWITHPNTPDKIYLHYABPTGPFVEPPDLFLLKTTKSKPLLGTTQMAKS 4653
QY      1009  PRKMTSTMPBLNPTSRAEAMLOTTTRPNOT---PNSKL-----1044
Db      4654  PKEMVSTPYKEAVLSGEBLDESSTSPNMSLQPLDQELSLSQPHGRIHPVTSQGIY 4713
QY      1045  ---VEYNPKS---EDAGGA 1057
Db      4714  LHYABPTGPFVEPPDLFLLKTTKAKPVQGTATHVEKSPKEMVFRSPYKQAVLPAHGEA 4773
QY      1058  EGET---PHMLL---RPHVEMPE-----VTPDMYL 1082
Db      4774  QDESPPNMSLQPLDQELTMSSQPHGWIPHPNTHGKIYLHYABPTGPFVEPPDLFLL 4833
QY      1083  PR-----VPOQ-----IINPM-----1095
Db      4834  KTTKSKPVQGTATQMAKSPPEMVSLLENKEAVLPAQDGLSEIFPPNMSLQDLHDLI 4893
QY      1096  LSDE-----TNI-----CNKGVDGLTUTTRNGTLV 1120
Db      4894  LSSQPHGWITHPNTPSNIVLHYABPTGPFVEPPDLFLLKTTKSKPVQ-----SPTLI 4948
QY      1121  A-----FRCH-----YFWMLSFSP 1135
Db      4949  AKSPREMSRSPYKQALLPGHGEKNESSAPNMSLQPLDQELTMSSQPYGW-----5001

```

```

QY 1136 FSPARRITEVMGIPSPIDTVTRCNCEGK-----TFFFKDSQ-----Y 1173
Db 5002 -----IHPHNT-----PCKIYLHYAEPPTGPFVEPPDLFFLTKTAKSPVQ 5042
QY 1174 WRFTNDIKDAGYKPIFKGFGGLTGQIVAAIATKAYK-----1210
Db 5043 W---TSABIAKLKPEIVS-----QTQKHESVLRAPEQHKNKSPSPENVSLQ 5086
QY 1211 -----NW-----PESVYFFKRGSGSIQYIYKQ 1232
Db 5087 PLDQELSLFSOPHGWIHPSPNTPGKIYLHYAEPPTGPFVEPPDLFFLK-----TTKS 5138
QY 1233 EPVQKCP-----GRPA-----1244
Db 5139 KPVQGTPTMTKSPQEIIVSQSPGMEAGPTTKVEQDESPPSHASLQTLDOELTLSSQP 5198
QY 1245 -----LNY-----PVYGMTQVRRRRFERAIG 1266
Db 5199 HGWIPHPNPTAGKIYLHYAEPPTGPFVEPPDLFFLTKTAKSPVQGTATQSTQ-----5250
QY 1267 PSQTHIRIQVSPARLAYQDKVLHNEVKVLSILWGLPNVVTSAI-----SLPNIRKP 1319
Db 5251 -----SPKEI-----ISPSTRYKESVLTAAVVGQDESQSPNLSLQ 5286
QY 1320 DGYDYAFSKDQYINIDVPSTRARAITRRSQTLKVM 1357
Db 5287 PLDQELTLSSQPHVLISHPSST-----NANTGHTLGLKIY 5320

RESULT 14
T25697
Hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GE/EMBL/DBRJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: UNIPROT:Q94185; EMBL:U67956; PIDN:AA807691.1; GSPDB:GN00028; CESP:F16F9
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP.F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.1%; Score 882.5; DB 2; Length 1229;
Best Local Similarity 23.7%; Pred.No. 4.8e-11;
Matches 375; Conservative 107; Mismatches 377; Indels 725; Gaps 79;

QY 2 AKNTLPIYLLLSVFIQVSSQDLSSCAGRCGGYSRATCNCYNCQHYMECCPDFK 61
Db 4 AW-VVSFAFLILGNVQSSLLSKTINSQSR-----DFK 37
QY 62 RV-----CTAELSCGRCFSEFGRGRCDDCAQCKYDKCCPQVSEFCAEVHNTSP- 113
Db 38 IVKHKKNTCT-----CSC-----KCPD-----APSNFF 62
QY 114 -----PSSKKAPPPSGAS-----OTIKSTYKSPKPNKKT 145
Db 63 DVSTTISSINNDVIGPSGDS-----NPTGSSWFQEIATVGGQTVKS-----E 107
QY 146 KKVISEEITEVKDNKNRKTKKKPKPPVDEAGSLDNGDFKVTTPDTSTTQ---HNK 202
Db 108 HNISSVEV-----EKKVT-----TSTDASTINAPTQK 136
QY 203 VSTSPKI-----TTAKPINRPSLPNPSDTSKTSLATVNKET 239
Db 1058 RA-----OPTN-----EMDKEMEFKIR 1076

```

```

Db 137 DSTTPEIITGIWVINSKSESVDMSSTRFSTTSLP-----TTPELLTSPT 181
QY 240 TVEIKETITTNKQSTDKKETTSAKETQSIKTSAXDLAPTCKVLAKPTPKAETTKGP 299
Db 182 LVSTDSSTST-EQTSFNDTTEIASPMET-----NTTE---213
QY 300 ALTPKPEPTPT---PKEPASTTPKEPTPTTIKAGPTTPKEP---APITTKSAPTTPKEP 353
Db 214 ATTTVSFVSTLASEDETIVTAIAESTTTVIAEVSTTTEBPTTTAESTTKKST---K 269
QY 354 APITTKPAPPTPKAPPTTTKAPPTTKSAPTTPKEPAPPTP---KKPAPTTPKEPA 409
Db 270 APATTEPTPTTEE---VTTEASTTSTISSETSEK---PTTLDINKIAGPATG---321
QY 410 PTTKPEP---TPTTPKPEAPTTPKEPAPTTP---KEPAPTAPKKPAPTTPKEPAPTTPKEPA 464
Db 322 PETTHFPVTGTP-----NFDATETPFVAKSEDKMTLSKTAATETTQOTTEV-DGPE 374
QY 465 PTTTKE-----PSPTTPKPEAPTTPKSGAP-----TTTKPEAPT---TTKSAPTTPKE 508
Db 375 KEITKNVSIPIITTVPLVETTTSTASKESDGFHTTLKLVTTADSDSTESATTVKPF 434
QY 509 PSPTTKE---PAPT-----TPK-----EPAPTTPKKPAP-----535
Db 435 NEETTKSHVVVPKTKGTUVKVPKLELSFDEPTEIT-KAPHGPKLLEKTYHFVLSDNF 493
QY 536 -----TTPKPEPA-----PTTPKPEAPT 552
Db 494 ARYSEAKENDDYNHLDYNHYREAKEPTTTEESSTTEBVTTEHPANTGNPTTEN---PT 550
QY 553 TTKPAPTAPKEA---PTTPKETAPTTPKGLTPTTPEKLAPTTPKEPAPTTPPELAPTTP 610
Db 551 TTEQPTSTABESTTALPFTTEQTV-----TTTE---PTTAEKSTATQ---KPTTT 594
QY 611 EEPPTTPPEPAPTTPKAAAPNTPKBPAPTTPKPEAPTTPKPEAPTTPKKTAP---TTPK 667
Db 595 QESVST---EKTSTTKA---STEE---PTTDEPTTTT---ESSTTGKATPELSSTSE 643
QY 668 GTAPTTLKEPAPTTPKKPAPKELAPTTPKEPTTSTGDKPAP---TTPKGTAPTTPKEPA 723
Db 644 ETTTTELK---ITTE-----GSTTTEPTTTFIAEASTGLITDETTSTTSTPE 692
QY 724 PTPKPEAPTTPKGTAPTTLKEPAPTTP-----KKPAPKELA 760
Db 693 ITSTKE---IVTESAITQTSVSVVESSTPRQLPERKAIIVNFKHNLVLEKRLKEKE 750
QY 761 PTTTKGP-----TSTTSKPAPTTPKETAPTTPKPEAPTTPKBPAPTTP 804
Db 751 STSTGSDSSETTVVAENIDEVTTTEKVKVQVTTTTEKSTTQEETTTTTTTTEKTT 810
QY 805 E---TPPPTTSE-VSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPGVPTTKP 861
Db 811 KUTTEKPTTSASATTTTSEPT---TEST-----TVDTS 843
QY 862 AATKPEMTTAKDKTTERDLRTTPETTA-----890
Db 844 SATTESSAETTTTSAE---TSETTTSESAAFITGESPENTALQSSQKSENESSAE 900
QY 891 -----APKMTKETATTTEKTTESKIFATTQO-VTST---TQDPTTPFKITLTKTTL 938
Db 901 KPGARDFVPKHKTTVPKPAETTSAAVASTTTEPITTEKSTTTEPITTEATTILNEVG 960
QY 939 APKVT---TTKKTITTTTBMKNPEETAKPKORATNSKATTPKQ-----979
Db 961 PAFVTGAPVDETTINTLELSK-----INNTOISQPKPTDISKTALSSILISGLI 1010
QY 980 KPTKAPKKPTSTKPKTMDPRVKPCTTPTPKMTSTMPELN-----PTSRIA 1026
Db 1011 GSFTKAPMPTI-----HTTTDAAFVATATEASLNDGSDKKIIDEAQPTDEIR 1057
QY 1027 EAMLQTTTTRPQTPNSKLVEYNPKSEDAAGAEFTPHMLLRPHVPMPEVTPMDYLRVP 1086
Db 1058 RA-----OPTN-----EMDKEMEFKIR 1076

```

QY 1087 NOGIII-----NPM----- 1096
 Db 1077 EQRIOEQAQLREELBELLKQOEIEBKARNEMIQRKMLQLEELKEAERQVILL 1136
 QY 1097 -----SDETNICNGKVPDGLTLRNLGTLVAFRGHYFWMSPSPSPAR- 1140
 Db 1137 EQRLEQERQRLIAEKEAIAFG-----SISITTEASKPKIR-----LRPAQC 1180
 QY 1141 ----RITVEYWGIPSIDTVFTRCNCE-GKTFKFDKSYWRFTNDIKDAGYKPKIFKGF 1195
 Db 1181 AAINKEFRVNTDPSEWI--QKNCEFAKYF-----PEASCP----- 1216
 QY 1196 LTGQIVAAALSTAKYKNWPESVYFF 1219
 Db 1217 ---QIQALI-----ESCFAP 1228
 RESULT 15
 B48666
 cell proliferation antigen Ki-67, short form - human
 C:Species: Homo sapiens (man)
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
 C:Accession: B48666
 R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
 J. Cell Biol. 123, 513-522, 1993
 A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
 ins.
 A:Reference number: A48666; MUID:94043435; PMID:8227122
 A:Accession: B48666
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2897 <SCH>
 A:Cross-references: EMBL:X65551
 C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
 F;29-91/Domain: kinase interaction domain homology <KIH>
 Query Match 12.0%; Score 881.4; DB 2; Length 2897;
 Best Local Similarity 14.1%; Pred. No. 3.2e-10;
 Matches 447; Conservative 151; Mismatches 457; Indels 2126; Gaps 111;
 QY 3 WKT-----LPVILLLSVFIQVSSQDLSSCAGRCGEGYSRDATNCND 47
 Db 2 WPTRLVTKRSGVDGHPFL-----SLSTCL----- 28
 QY 48 YNCQHYMECCPDFKRVCTAELSCKRCFESFERGECDDAQ-----CKKYDKCCPD 99
 Db 29 -----FGRGIECDIRIQLPWSKQHC----- 50
 QY 100 YESFCAEVH-----NPTS----- 112
 Db 51 -----IEIHEQEAHLNFSSTNPQVNGSVIDEVRLKHGDVITIDRSFYENESLQNG 105
 QY 113 -----PPS 115
 Db 106 RKSTFPRKIREQPARVRSRSPSSDDEKAQDSKAYSKITEGKVGSNPELFDENLPPN 165
 QY 116 S-----KKAPPPSG----- 124
 Db 166 TPLRGEAPTKRKSLVMHTPPVLKKIIEQOPQSGKQSGSEIHVEVKAQSLIVISPPAPS 225
 QY 125 -----ASQITKSTTKRSPKPNKKTK----- 146
 Db 226 PRKTPVASDQRRRSCKTAPASSSKSQTEVPKRGGERVATCLQKRVISRSRSHDILQMTCS 285
 QY 147 -----KVIESBEITEVKNKNRTKKKPK-TPKPPV----- 175
 Db 286 KRRGASEANLIVAKSNADVVKLGAKQQTQVKVHKGPQRSMMKRRRPAATPKKPVGEVHS 345
 QY 176 -----VDEAG----- 180
 Db 346 QFSTGHANSPTIILIGKAHTEKVVHVPARYVRLNNFISNQKMDFKEDLSGAEAFKTPVK 405

QY 181 -----SGLDNGD----- 187
 Db 406 EQPQLTSTCHIALSENSENLLGKQFQGTDSGEPLLPSTSEFGNVFFSAQNAAKQPSDKC 465
 QY 188 -----FKVT-----TPDTSTTOHNKYST-----SPALITAKPI 215
 Db 466 SASPLRROCIRENGNVAKTPRNTYKMTSLETNTKSTETETPKTVSTVARSGRSTFRNI 525
 QY 216 NRP----- 219
 Db 526 QKLPVESKSEETNTEIVECILKRGQXATLLQORREGEMKEIERPFTYKENIELKENDEK 585
 QY 220 -----SLP-----PNSDTSKETS 232
 Db 586 MKAMKESRIWQKCAPMSDLTDLKSLPDTLMKDTARGQNLQTQDHAKAPKSEKGIKTX 645
 QY 233 LTVNK-----ETTVETK-----TTTTNKQTSDDGKE 259
 Db 646 MPCQSLQPEPINTPHTTKQQLKASLGKVGKVEELLAVGKFTRTSGETTHRBPAGDGKS 705
 QY 260 -----KTTSAKETQSIEKTSAKDLAPTSKVLAKTPPKAET- 294
 Db 706 IRTFKESPQILDPAARVTCMKKWPRTPEEAQSL-----DLAGPKELFTQTPSPSESM 760
 QY 295 -----TTK-----GPAITTPKE 306
 Db 761 TDEKTTKIACKSPPPESVDPTSTKQWPKRSLKADVEEBFLALRLKLTSPSAGKAMLTTPK 820
 QY 307 P-----TPT-----TPKEPAST----- 318
 Db 821 AGGDEKDIKAFMGTPVQKLDLGLTFSGSKQLQTPKEKAQALDGLAGKELFQTPGHTEE 880
 QY 319 -----TPKEP-----TPTTKSAPT-----TP 335
 Db 881 LVAAGTKTKPCDSQSDPDVDTSTKQPKRSIRKADVEGELLACRLNMPKSAGKAMHTP 940
 QY 336 KEPAP-----TTTSAPTTPKEPAPT----- 356
 Db 941 K---PSVGEBKDIIIFVGTVPQKLDLTENLTGSKRRRQTPKEAQALEDLTGFKELFQTP 997
 QY 357 -----TKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKE----- 391
 Db 998 GHTEEAAGAAGTKTKMPCESSPPESADT-----PTSTRQPKTFLEKRDVQKELSAUKL 1051
 QY 392 -----PAPTTPKPAPTTPKEPA----- 409
 Db 1052 TQTSGETHTDKVPGGEDKSINAFRETAQKQLDPAASVTGSKRHPT-KEKAQPLEDLAG 1110
 QY 410 -----PTTPK-----EPPT-TPKEPAPTKE----- 430
 Db 1111 WKELFQTPVCTDKPTTHEKTTKIACRSQDPDPTSTSSKQSKSLRKVDVEEBFFALRK 1170
 QY 431 -----PAPT-----TPKEPA----- 440
 Db 1171 RTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLTENLTGSKRRRLQTPKEKAQALD 1230
 QY 441 -----PTAPKKA----- 448
 Db 1231 GFKELFQTRGHTEESMTNDKTAKVACKSSQDLDKNPASSKRRRLKTSLGKVGKVEELLAV 1290
 QY 449 -----PTTPKBPAPT-----TPKEPA----- 464
 Db 1291 GKLQTSGETHTHTEPTGDKGSKMAKPMESPKQLDASAISLTGSKRQLRTPKGSKEVPD 1350
 QY 465 -----PTTKESPTTPK-----EP-----APTTSKAP----- 488
 Db 1351 LAGFIELFQTPSHTKB-SMTNEKTKVSYRASQDPLDVTPTSSKQPKRSRKADTEEF 1409
 QY 489 -----TTTKEPA----- 495
 Db 1410 LAFKQTPSAGKAMHTPKPAVGEKQINTFLGTVPQKLDQGNLPGSNRRRLQTRKEKAQ 1469
 QY 496 -----PT-----TTK-----SAPTTPKEPSPTTKE----- 516

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 105.696 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-H

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKVWVNCV 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7058	98.9	1311	8	Adk67912 Human ext
2	7044.7	99.7	1404	2	Aar26049 MSF precu
3	7044.7	99.7	1404	4	Aab29773 Human meg
4	7044.7	99.7	1404	4	Aab60568 Human meg
5	7044.7	99.7	1404	8	Adm98014 Human meg
6	7041.7	99.7	1415	4	Aau32262 Novel hum
7	7029.7	99.5	1404	7	Adk65839 Angiogene
8	6793.9	96.2	1270	8	Adk67911 Human ext
9	6769.9	95.8	1320	7	Adk65819 Angiogene
10	6468.7	91.6	1299	4	Aam24322 Human ESI
11	3552	50.3	902	4	Aab29778 Human MSF
12	2929	41.5	551	4	Abu53253 Human tes
13	2820	41.3	546	4	Abu53252 Human tes
14	2850.9	40.4	538	5	Aao18834 3' cartil
15	2757	39.0	513	4	Abu53254 Human tes
16	2009	28.4	472	4	Aab60569 Bovine MS
17	1958.9	27.7	452	2	Aar80041 Human meg
18	1398.2	19.8	5179	4	Aam24516 C899p pre
19	1398.2	19.8	5179	6	Abp55365 Human col
20	1398.2	19.8	5179	6	Abc07258 Human p53
21	1398.2	19.8	5179	7	Add48091 Human Pro
22	1398.2	19.8	5179	7	Add44998 Human Pro
23	1307.1	18.5	292	5	Aau11261 Human HAP
24	1154.2	16.3	1664	2	Aaw43106 C. thermo
25	1139.7	16.1	8991	6	Abu08487 S. pneumo

26	1112.2	15.7	5703	8	ADL23265	Human MUC
27	1099.6	15.6	717	4	ABU53144	Human tes
28	1097	15.5	214	4	ABU53255	Human tes
29	1096.4	15.5	1795	4	ABB69806	Drosophil
30	1082.8	15.3	763	3	RAG38942	Drosophil
31	1073.5	15.2	4315	5	ABP43908	Arabidops
32	1054.5	14.9	2284	4	ABB71434	MUC5B par
33	1039.7	14.7	1049	4	ABB61364	Drosophil
34	1025	14.5	188	5	AAO18833	Abb61364
35	1001.8	14.2	4498	4	ABB58595	Drosophil
36	992.7	14.1	770	4	ABU53141	Drosophil
37	987.6	14.0	778	4	ABU53143	Human tes
38	974.8	13.8	2112	4	ABB60403	Human tes
39	974.4	13.8	692	4	ABU53155	Drosophil
40	969.9	13.7	1460	7	ADN39110	Human tes
41	969.6	13.7	1538	7	ADI21202	Cancer/an
42	967.2	13.7	2768	4	ABB68397	Novel hum
43	964.1	13.6	717	4	ABU53150	Drosophil
44	964.1	13.6	717	4	ABU53149	Human tes
45	964.1	13.6	717	4	ABU53151	Human tes

ALIGNMENTS

RESULT 1

ADK67912

ID ADK67912 standard; protein; 1311 AA.

AC ADK67912;

DT 06-MAY-2004 (first entry)

XX Human extracellular messenger (EXMBS) polypeptide.

DE Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;

KW antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;

KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;

KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;

KW cytostatic; gene therapy.

XX Homo sapiens.

OS Homo sapiens.

XX Key

PH Location/Qualifiers

FT Peptide

FT 1..30

FT /label= Signal peptide

FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30

FT according to identification method"

XX WO2004013292-A2.

XX 12-FEB-2004.

XX 30-JUL-2003; 2003WO-US024084.

XX 02-AUG-2002; 2002US-0400810P.

XX 19-SEP-2002; 2002US-0412197P.

XX 04-OCT-2002; 2002US-0416004P.

XX 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

XX Richardson TW, Emerling BM, Lindquist EA, Ramkumar J;

XX Lee SY;

XX WPI; 2004-157116/15.

XX N-FSDB; ADK67917.

XX New extracellular messengers and nucleic acids, useful for diagnosing,

XX treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes

XX mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or

XX autoimmune thyroiditis.

XX	Claim 60; SEQ ID NO 5; 165pp; English.
PS	The present sequence is that of novel human extracellular messenger
CC	(EXMES) incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
CC	to human megakaryocyte stimulating factor. The invention provides EXMES
CC	polynucleotides and polypeptides, as well as expression vectors, host
CC	cells, antibodies, agonists and antagonists, and methods for diagnosing,
CC	treating or preventing disorders associated with aberrant expression of,
CC	EXMES, especially autoimmune and inflammatory disorders, e.g. adult respiratory
CC	proliferative disorders and endocrine disorders, e.g. foetal Crohn's
CC	disease syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
CC	disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
CC	osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
CC	parasitic, protozoal or helminthic infections, cancers, autoimmune
CC	thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
CC	Embodiments also provide methods for using the purified EXMES and/or
CC	their encoding polynucleotides for facilitating the drug discovery
CC	process, including determining of efficacy, dosage, toxicity and
CC	pharmacology, and for investigating the pathogenesis of diseases and
CC	medical conditions.
XX	
SQ	Sequence 1311 AA;
	Query Match 99.9%; Score 7058; DB 8; Length 1311;
	Best Local Similarity 99.9%; Pred. No. 3.4e-171;
	Matches 1310; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 MAWKTLPIYLLLLLVFIQQVSSODLSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
DB	
Db	1 MAWKTLPIYLLLLLVFIQQVSSODLSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
QY	61 KRVCCTAELSCKGRCFESFERGERECDDAQCKKYDKCCPDYESFCAEVNDKNRNTKKKPT 120
Db	
Db	61 KRVCCTAELSCKGRCFESFERGERECDDAQCKKYDKCCPDYESFCAEVNDKNRNTKKKPT 120
QY	121 PKPPVVDEAGSGLDNGDFKVTTDPDTSITQHNVKSVTSPKLTITAKPINRPSPNPNSDTSKE 180
QY	
Db	121 PKPPVVDEAGSGLDNGDFKVTTDPDTSITQHNVKSVTSPKLTITAKPINRPSPNPNSDTSKE 180
QY	181 TSLTVNKETTVTKETTTNNKOTSIDGKEKTTSAKETQSIEKTSADKLAPTSKVLAQTP 240
Db	
Db	181 TSLTVNKETTVTKETTTNNKOTSIDGKEKTTSAKETQSIEKTSADKLAPTSKVLAQTP 240
QY	241 XAETTITKGPAITTPKEBPTPTTPKEPASTTPKGPPTTTKSAPTTPKEPATPTTKSAPTP 300
Db	
Db	241 XAETTITKGPAITTPKEBPTPTTPKEPASTTPKGPPTTTIISAPTTPKEPATPTTKSAPTP 300
QY	301 KBPAPTTTKEPATTTKEPATTTKEPATTTKSAPTTPKEPATTPPKKPAPTTPKECAP 360
Db	
Db	301 KESAPTTPKEPATTPKEPATTTKEPATTTTTSAPTTPKEPATTPKKEPATTPKECAP 360
QY	361 TTPEKEPTPTTPKEPATTTKEPATTTKEPATTAPEKPAATTPKEPATTPKEPATTTTKE 420
Db	
Db	361 TTPEKEPTPTTPKEPATTTKEPATTTKEPATTAPEKPAATTPKEPATTTTKEPATTTTKE 420
QY	421 PSSTTPKEPATTTKSAPTTPKEPATTTKSAPTTPKEPSPTTTPKEPATTTTKEPATPTP 480
Db	
Db	421 PSSTTPKEPATTTKSAPTTPKEPATTTKSAPTTPKEPSPTTTPKEPATTTTKEPATPTP 480
QY	481 KKCAPITTPKEPATTTKEPATTTTCKPAPTAPEKPAATTPKEPATTTPKKLTPTTPEKIA 540
Db	
Db	481 KKCAPITTPKEPATTTTKEPATTTTCKPAPTAPEKPAATTPKEPATTTPKKLTPTTPEKIA 540
QY	541 PTTPEKAPPTTPEBLAPTTPBEPTPTTPEEPATTTKAAANTPKEPATTTTKEPATPTP 600
QY	
Db	541 PTTPEKAPPTTPEBLAPTTPBEPTPTTPEEPATTTKAAANTPKEPATTTTKEPATPTP 600
QY	601 KEPAPTTPKETAPTTPKGTAPTTLKEPATTTTTPKKPAPEKELAPTTTTKEPTSTTSDKAPPT 660
QY	
Db	601 KEPAPTTPKETAPTTPKGTAPTTLKEPATTTTTPKKPAPEKELAPTTTTKEPTSTTSDKAPPT 660
QY	661 PKGAPTTPKEPATTTTKEPATTPTKGAPTTLKEPATTTTTPKKPAPEKELAPTTTTKEPTST 720

FT Region 1166..1212
 FT /label= Exon_VIII
 FT Region 1213..1266
 FT /label= Exon_IX
 FT Region 1266..1331
 FT /label= Exon_X
 FT Region 1331..1373
 FT /label= Exon_XI
 FT Region 1373..1404
 FT /label= Exon_XII
 FT Region 1411..1166
 FT /label= Exon_VII
 PN WO9213075-A1.
 XX
 PD 06-AUG-1992.
 XX
 PF 17-JAN-1992; 92WO-US000433.
 XX
 PR 18-JAN-1991; 91US-00643502.
 PR 10-SEP-1991; 91US-00757022.
 XX
 PA (GEM) GENETICS INST INC.
 XX
 PI Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;
 XX
 DR WPI; 1992-284660/34.
 DR N-PSDB; AAQ27223.
 XX
 PT New human mega-karyocyte stimulating factors - for treating immune
 PT deficiencies, cancer, exposure to radiation or drugs, bacterial and viral
 PT infections, etc.
 PS
 PS Claim 1, 2 and 3; Fig 1; 87pp; English.
 XX
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II, III
 CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is
 CC modified by the addition of an N-terminal sequence encoding a secretory
 CC leader, an initiating methionine preceding exon II and a terminating
 CC codon following exon IV. The cDNA sequence given contains sequences
 CC derived from human megakaryocyte colony stimulating factor (meg-CSF).
 CC Exon I contains the initiating methionine, and encodes a classical
 CC mammalian protein secretion signal sequence. The sequence encoding the
 CC original meg-CSF includes exons II-IV and is thought to terminate in the
 CC region between amino acid residues 134 - 147. The primary transcript of
 CC this gene may be cleaved in different ways to yield a family of mRNA's
 CC each encoding a different MSF protein. Exons V and VI are thought to be
 CC related to the activity of the factor and are also implicated in the
 CC stability, folding and processing of the molecule. These exons are also
 CC thought to play a role in the observed synergy of MSF with other
 CC cytokines. Exons V - XII are believed to be implicated in the processing
 CC or folding of the appropriate structure of the resulting factor, ie. one
 CC or more of these exons may contain sequences which direct proteolytic
 CC cleavage, adhesion, organisation of the cellular matrix or extracellular
 CC matrix processing. Both naturally occurring and non-naturally occurring
 CC MSF's may be characterised by various combinations of alternatively
 CC spliced exons from this sequence, with the exons spliced together in
 CC differing orders to form different members of the MSF family. (Updated on
 CC 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 1404 AA;
 Query Match 99.7%; Score 7044.7; DB 2; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 8.2e-171;
 Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
 QY 1 MAWKTLPIYLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCOHYMECCPDF 60
 DB 1 MAWKTLPIYLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCOHYMECCPDF 60
 QY 61 KRVCTAELSCGRCFESFERGECDDAQCCKYDKCCPDYSEFCAE----- 106
 |||||||

Db 61 KRVCTAELSCGRCFESFERGECDDAQCCKYDKCCPDYSEFCAE VHNPTSPSSKKAP 120
 QY 107 ----- 106
 Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVBIESEEITEBHSYSENOESSSSSSSSSTI W 180
 QY 107 -----VKDNKNKNTKKPKPPVVDVDEAGSLGNGDFKVTTPDTST 147
 Db 181 KIKSSKNSAANRELQKKLVKDNKNKNTKKPKPPVVDVDEAGSLGNGDFKVTTPDTST 240
 QY 148 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTVEVTKETITNKQTSIDG 207
 Db 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTVEVTKETITNKQTSIDG 300
 QY 208 KEKTSIAKETQSIEKTSAKDLAPTSKVLAKPPTKPAETTTKGPALTTTPEKPTTTPKEPAS 267
 Db 301 KEKTSIAKETQSIEKTSAKDLAPTSKVLAKPPTKPAETTTKGPALTTTPEKPTTTPKEPAS 360
 QY 268 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTKPAPTTTKP 327
 Db 361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTKPAPTTTKP 420
 QY 328 APTTIKSAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKP 387
 Db 421 APTTIKSAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKP 480
 QY 388 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKPAPTTTKSAPTTTKPAPT 447
 Db 481 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKPAPTTTKSAPTTTKPAPT 540
 QY 448 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTKKP 507
 Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKEPAPTTTKKP 600
 QY 508 APTAPKEPAPTTTPKETAPTTTPKKLTPTTPEKLAAPTTPKEPAPTTPELAPTTPEEPTPTT 567
 Db 601 APTAPKEPAPTTTPKETAPTTTPKKLTPTTPEKLAAPTTPKEPAPTTPELAPTTPEEPTPTT 660
 QY 568 PEEPAPTTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTLKEP 627
 Db 661 PEEPAPTTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTLKEP 720
 QY 628 APTTPKAPKELAPTTTIKEPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKGPAPTTPKG 687
 Db 721 APTTPKAPKELAPTTTIKEPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKGPAPTTPKG 780
 QY 688 TAPTTLKEPAPTTTPKAPKELAPTTTIKGTPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPK 747
 Db 781 TAPTTLKEPAPTTTPKAPKELAPTTTIKGTPTSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPK 840
 QY 748 KPAPTTTPPTTSEVSTPTTIKEPTTIHKSPESTPELSAEPPTKALENSPKPGVPT 807
 Db 841 KPAPTTTPPTTSEVSTPTTIKEPTTIHKSPESTPELSAEPPTKALENSPKPGVPT 900
 QY 808 TKTPAATPEMTTAKDKTTREDLNTTPTTAAAPKMTKETATTTKTESKITATTQV 867
 Db 901 TKTPAATPEMTTAKDKTTREDLNTTPTTAAAPKMTKETATTTKTESKITATTQV 960
 QY 868 TSTTTQDTPPTTKITLTKTTLLAPKVTTTKITLTTTINMKPEETAAPKDRATNSKATTPK 927
 Db 961 TSTTTQDTPPTTKITLTKTTLLAPKVTTTKITLTTTINMKPEETAAPKDRATNSKATTPK 1020
 QY 928 PQKPTAPKPTSTTKPKIMPRVRKPKTTTTPRKWNTSTMPELNPTSRIAEAMLOTTTPRN 987
 Db 1021 PQKPTAPKPTSTTKPKIMPRVRKPKTTTTPRKWNTSTMPELNPTSRIAEAMLOTTTPRN 1080
 QY 988 QTPNSKLVENVNPKSEDAGGAEGETPHMLLRHVFMPEVTPDMDYLPVNPNGIINPMLS 1047
 Db 1081 QTPNSKLVENVNPKSEDAGGAEGETPHMLLRHVFMPEVTPDMDYLPVNPNGIINPMLS 1140
 QY 1048 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPEFPPSPARRITTEVWGTPSIDTIVFT 1107
 Db 1141 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPEFPPSPARRITTEVWGTPSIDTIVFT 1200

QY 1108 RNCCEGKTFKFDKSYWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESY 1167
DB 1201 RNCCEGKTFKFDKSYWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESY 1260
QY 1168 FPKRGSGTQOYIYKQEPVQKCPGRPALNPVYVGEVQVRRRFRRAIGPSQTHIRIQY 1227
DB 1261 FPKRGSGTQOYIYKQEPVQKCPGRPALNPVYVGEVQVRRRFRRAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
DB 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
QY 1288 PSTRARAITRSGQTLKQVWYNCP 1311
DB 1381 PSTRARAITRSGQTLKQVWYNCP 1404

RESULT 3
AAB29773
ID AAB29773 standard; protein; 1404 AA.
XX AAB29773;
XX 28-FEB-2001 (first entry)
DT Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
XX Human MSF; megakaryocyte stimulating factor; tribonectin;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX Homo sapiens.
OS W0200064930-A2.
PN 02-NOV-2000.
XX 24-APR-2000; 2000MO-US010953.
XX 23-APR-1999; 99US-00298970.
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX Jay GD;
XX WPI; 2001-024673/03.
XX N-PSDB; AAC81498.
XX Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
PS Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonectin; a biocompatible composition comprising a human
CC tribonectin for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
CC used in gene therapy. The present sequence represents human MSF
XX Sequence 1404 AA;
SQ Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 8.2e-171;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KVCCTAELSCKRCFESFERGECDCDAOCKKYDKCCPDYEFCAE----- 106
DB 61 KVCCTAELSCKRCFESFERGECDCDAOCKKYDKCCPDYEFCAE----- 120
QY 107 ----- 106
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIJESBEITEHSVSENQESSSSSSSSSSTIW 180
QY 107 -----VKONKQRTKKKPTKPPVVDGAGSLDNGDFKVTDTST 147
DB 181 KIKSSKNSAANRELQKLVKDKNRTKKKTPKPPVVDGAGSLDNGDFKVTDTST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKETTNNKQSTDG 207
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKETTNNKQSTDG 300
QY 208 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKGPPTTPKPEPAS 267
DB 301 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKGPPTTPKPEPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
DB 361 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY 328 APTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 387
DB 421 APTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 480
QY 388 EPAPTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKPSPTTPKPEAPTTTKSAPTTPKPEAPT 447
DB 481 EPAPTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKPSPTTPKPEAPTTTKSAPTTPKPEAPT 540
QY 448 TTKSAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 507
DB 541 TTKSAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 600
QY 508 APTAPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 567
DB 601 APTAPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 660
QY 568 PEPAPPTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 627
DB 661 PEPAPPTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPT 720
QY 628 APTTPKPEAPKELAPTTTKPEPTSTSDKPARTTKGTAPTTPKPEAPTTTPKPEAPTTPKG 687
DB 721 APTTPKPEAPKELAPTTTKPEPTSTSDKPARTTKGTAPTTPKPEAPTTTPKPEAPTTPKG 780
QY 688 TAPTTLKEAPPTPKPEAPKELAPTTTKGPTSTSDKPARTPKETAPTTPKETAPTTPKETAPTT 747
DB 781 TAPTTLKEAPPTPKPEAPKELAPTTTKGPTSTSDKPARTPKETAPTTPKETAPTTPKETAPTT 840
QY 748 KPAPTTPETPTPTSEVSTPTTKPEPTTIHKSPEDESTPELSAETPKALENSPKERGVP 807
DB 841 KPAPTTPETPTPTSEVSTPTTKPEPTTIHKSPEDESTPELSAETPKALENSPKERGVP 900
QY 808 TKTAATKPEMTTAKDKTTERDURTPETTTAAKPKTKETATTTKTESKITATTTQV 867
DB 901 TKTAATKPEMTTAKDKTTERDURTPETTTAAKPKTKETATTTKTESKITATTTQV 960

QY 868 TSTTTQDTPPKITLTKITLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATPK 927
 Db 961 TSTTTQDTPPKITLTKITLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATPK 1020
 QY 928 POKETKAPKPTSTKPKTMRVRKPKTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN 987
 Db 1021 POKETKAPKPTSTKPKTMRVRKPKTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN 1080
 QY 988 QTPNSKLVEVNPKESEDAGGEGETHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1047
 Db 1081 QTPNSKLVEVNPKESEDAGGEGETHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1140
 QY 1048 DETNINCNGKPDVGLTLTLRNGTLVAFRGHYFWMLSPFPSPARRITVWGPSPIDIVET 1107
 Db 1141 DETNINCNGKPDVGLTLTLRNGTLVAFRGHYFWMLSPFPSPARRITVWGPSPIDIVET 1200
 QY 1108 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYNWPEVY 1167
 Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYNWPEVY 1260
 QY 1168 FFKRGGSIOQYIYKQBPVKPCGRRRPALNYPVYGEVTOVRRRPFERRAIGPSQTHIRIQY 1227
 Db 1261 FFKRGGSIOQYIYKQBPVKPCGRRRPALNYPVYGEVTOVRRRPFERRAIGPSQTHIRIQY 1320
 QY 1228 SPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYNNIDV 1287
 Db 1321 SPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYNNIDV 1380
 QY 1288 PSRTARAITRRSQTLTKWYNCP 1311
 Db 1381 PSRTARAITRRSQTLTKWYNCP 1404

RESULT 4

AAB60568
 ID AAB60568 standard; protein; 1404 AA.
 XX AAB60568;
 AC AAB60568;
 DT 27-APR-2001 (first entry)
 XX Human megakaryocyte stimulating factor (MSF, CACP).
 DE Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KW antiarthritic.
 XX Homo sapiens.
 OS WO200107068-A1.
 PN 01-FEB-2001.
 PD 21-JUL-2000; 2000WO-US020002.
 PF 23-JUL-1999; 99US-0145328P.
 PR 19-JUL-2000; 2000US-00145328.
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 PA Warman ML;
 PI WPI; 2001-182721/18.
 DR New composition comprising the camptodactyly-arthropathy-coxa vara-
 PT pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints.
 XX Example 1; Page; 34pp; English.
 PS The invention relates to a method of treating osteoarthritis via the

administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)

XX Sequence 1404 AA;

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 8.2e-171; Indels 93; Gaps 1;
 Matches 1311; Conservative 0; Mismatches 0;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
 Db 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
 QY 61 KRVTAEALSCGRCFESFERGECDDAOCKKYDKCCPDYESFCAE----- 106
 Db 61 KRVTAEALSCGRCFESFERGECDDAOCKKYDKCCPDYESFCAE----- 120
 QY 107 ----- 106
 Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESBEITEHSVSENQESSSSSSSSSSSTIW 180
 QY 107 -----VKONKKNRKTKKTPKPPVVDEAGSLDNGDFKVTTPDTST 147
 Db 181 KIKSSKNSAANRELQKKLVKDKNKNRKTCKKTPKPPVVDEAGSLDNGDFKVTTPDTST 240
 QY 148 TQHNKVTSPKLTAKPINPRESLPNSDTSKETSITVKNKETTIVETKETTNNKQSTDG 207
 Db 241 TQHNKVTSPKLTAKPINPRESLPNSDTSKETSITVKNKETTIVETKETTNNKQSTDG 300
 QY 208 KEKTTSAKETQSIEKTSADLAPTSKVLAKPDKAETTTKGALTTTPKEPTTTPKEPAS 267
 Db 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPDKAETTTKGALTTTPKEPTTTPKEPAS 360
 QY 268 TTPKEPTTTPKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP 327
 Db 361 TTPKEPTTTPKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP 420
 QY 328 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 387
 Db 421 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 480
 QY 388 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTTTPKEPAP 447
 Db 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTTTPKEPAP 540
 QY 448 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKP 507
 Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKP 600
 QY 508 APTAPKEPAPTTTPKETAPTTTPKLTTPTEKLAOTTPKEPAPTTTPKEPAPTTTPKEP 567
 Db 601 APTAPKEPAPTTTPKETAPTTTPKLTTPTEKLAOTTPKEPAPTTTPKEPAPTTTPKEP 660
 QY 568 PEEPAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 627
 Db 661 PEEPAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 720

QY 628 APTPKKAPKELAPTTTKETSTSDKPAPTTPKGTAPTTKPAPTTPKEPAPTTKPKG 687
DB 721 APTPKKAPKELAPTTTKETSTSDKPAPTTPKGTAPTTKPAPTTPKEPAPTTKPKG 780
QY 688 TAPTTLKPAPTTPKPAKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPAPTTKPK 747
DB 781 TAPTTLKPAPTTPKPAKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPAPTTKPK 840
QY 748 KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALENSPKPGVPT 807
DB 841 KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALENSPKPGVPT 900
QY 808 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAKMTKETATTTKTESKITATTQV 867
DB 901 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAKMTKETATTTKTESKITATTQV 960
QY 868 TSTTTQDTTPFKITLTKTTLAPKVTITTKTITTEIMNKPETAKPKDRATNSKATTPK 927
DB 961 TSTTTQDTTPFKITLTKTTLAPKVTITTKTITTEIMNKPETAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKKPTSTKKPKTMPVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRN 987
DB 1021 POKPTKAPKKPTSTKKPKTMPVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRN 1080
QY 988 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMYLRVNPQGIINPMLS 1047
DB 1081 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMYLRVNPQGIINPMLS 1140
QY 1048 DETNINCGPVDGLTTLRNGTLVAFRGHYFWMLSPTSPSPARRITEVWGIPSPIDTFT 1107
DB 1141 DETNINCGPVDGLTTLRNGTLVAFRGHYFWMLSPTSPSPARRITEVWGIPSPIDTFT 1200
QY 1108 RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAISTAKYKNWPEVY 1167
DB 1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAISTAKYKNWPEVY 1260
QY 1168 FFRGGSIOQYIYKQEPVQKCPGRPALNYPVYGEWTOVRRRPERAIGPSQHTIRIQY 1227
DB 1261 FFRGGSIOQYIYKQEPVQKCPGRPALNYPVYGEWTOVRRRPERAIGPSQHTIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDCGYDYAFSKQYYNIDV 1287
DB 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDCGYDYAFSKQYYNIDV 1380
QY 1288 PSRTARAITRSGOTLSKWYNCP 1311
DB 1381 PSRTARAITRSGOTLSKWYNCP 1404

RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.

XX AC ADM98014;

XX 01-JUL-2004 (first entry)

DE Human megakaryocyte stimulating factor (MSF).

XX lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.

XX OS Homo sapiens.

XX PN US2004072741-A1.

XX 15-APR-2004.

XX 02-JUL-2001; 2001US-00897188.

XX 23-APR-1999; 99US-00298970.

PR 24-APR-2000; 2000US-00556246.

XX (JAYG/) JAY G D.
PA Jay GD;
PI WPI: 2004-373948/35.
DR N-PSDB; ADM98015.
XX New tribonectin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
XX Claim 1; SEQ ID NO 1; 34pp; English.
XX The invention relates to a lubricating polypeptide and at least one O-
CC linked oligosaccharide. The composition and methods are useful for
CC lubricating joints or other tissues to prevent or treat camptodactyl-
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
CC sequence represents the amino acid sequence of the human megakaryocyte
CC stimulating factor (MSF).

XX Sequence 1404 AA;

Query Match 99.7%; Score 7044.7; DB 8; Length 1404;
Best Local Similarity 93.4%; Pred. No. 8.2e-171;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGCGEGYGRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGCGEGYGRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTAEELCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- 106
DB 61 KRVTAEELCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- 120
QY 107 ----- 106

DB 121 PPSGASQTIKSTTRGPKPPNKKTKKVIIESEBITBEHSVSENQESSSSSSSSSSSTIW 180
QY 107 -----VKONKMKRTKKPKPPVVDAGSGLDNGFKVTTPTDTST 147
DB 181 KIKSSKNANRELQKKLVKONKMKRTKKPKPPVVDAGSGLDNGFKVTTPTDTST 240
QY 148 TQNKVSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKEITVETKETTNNKQSTDG 207
DB 241 TQNKVSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKEITVETKETTNNKQSTDG 300
QY 208 KEKTSAKETQSIETKSADLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPPKEPAS 267
DB 301 KEKTSAKETQSIETKSADLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPPKEPAS 360
QY 268 TTPKEPTTPTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 327
DB 361 TTPKEPTTPTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 420
QY 328 APITTKSAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPTTPPKEPAPTTTPK 387
DB 421 APITTKSAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPTTPPKEPAPTTTPK 480
QY 388 EPAPTAAPKAPPTTPKEPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTTKEPAPT 447
DB 481 EPAPTAAPKAPPTTPKEPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTTKEPAPT 540
QY 448 TTKSAPTTTPKEPSPTTTPKEPAPTTTPKAPTTTPKAPTTTPKEPAPTTTPKEPAPTTT 507
DB 541 TTKSAPTTTPKEPSPTTTPKEPAPTTTPKAPTTTPKAPTTTPKEPAPTTTPKEPAPTTT 600
QY 508 APTAPKPEPAPTTTPKETAPTTPKLTPPTPEKLAFTTPEKAPTTPEELAPTTPEEPTPT 567
DB 601 APTAPKPEPAPTTTPKETAPTTPKLTPPTPEKLAFTTPEKAPTTPEELAPTTPEEPTPT 660
QY 568 PEEPAPTTPKAAANTPKAPPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTATPTLKEP 627

Db	661	PEEPAPTPKAAAPNTPEKAPATTPKEPAPTPKAPATTPKETAATTPKGTATPTTLKEP	720
Qy	628	APTTPKPAKELAPATTTKBTSTSDKAPATTTKGTAPATTPKAPATTPKGPATTPKG	687
Db	721	APTTPKPAKELAPATTTKBTSTSDKAPATTTKGTAPATTPKGPATTPKGPATTPKG	780
Qy	698	TAPTTLKEPAPTPPKAPKELAPATTTKGTSTSDKAPATTPKETAATTPKGPATTPK	747
Db	781	TAPTTLKEPAPTPPKAPKELAPATTTKGTSTSDKAPATTPKETAATTPKGPATTPK	840
Qy	748	KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSABPTPKALENSPKPCVPT	807
Db	841	KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSABPTPKALENSPKPCVPT	900
Qy	808	TKTPAATPKPMWTTAKDKTTERDLRTTPEPTTAAPKMTKETATTTKTTESKITTATTVQ	867
Db	901	TKTPAATPKPMWTTAKDKTTERDLRTTPEPTTAAPKMTKETATTTKTTESKITTATTVQ	960
Qy	868	TSTTTQDTPFKITTLTKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK	927
Db	961	TSTTTQDTPFKITTLTKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK	1020
Qy	928	POKPTAPKPKPTSTKPKMVRKPKTTPTPRKMSTWPELNPSTRIAEAMLQTTTRPN	987
Db	1021	POKPTAPKPKPTSTKPKMVRKPKTTPTPRKMSTWPELNPSTRIAEAMLQTTTRPN	1080
Qy	988	QTPNSKLVEYNPKSEDAGAGETPEMLLRPHVFMPEVTPDMDLPRVNOGIIINPMLS	1047
Db	1081	QTPNSKLVEYNPKSEDAGAGETPEMLLRPHVFMPEVTPDMDLPRVNOGIIINPMLS	1140
Qy	1048	DETNIENGVPDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVVGIPSPIDTVFT	1107
Db	1141	DETNIENGVPDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVVGIPSPIDTVFT	1200
Qy	1108	RNCEGKTFEPKDSQWRFETNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYNWPEVY	1167
Db	1201	RNCEGKTFEPKDSQWRFETNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYNWPEVY	1260
Qy	1168	FFKRGSSIQQYIYKQEPVOKCGRPALNVPVVGEMTQVRRRRFRRAIGPSQTHIRIQY	1227
Db	1261	FFKRGSSIQQYIYKQEPVOKCGRPALNVPVVGEMTQVRRRRFRRAIGPSQTHIRIQY	1320
Qy	1228	SPARLAYQKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYVNDV	1287
Db	1321	SPARLAYQKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYVNDV	1380
Qy	1288	PSRTARAITTRSGQTLKSWWYNCP	1311
Db	1381	PSRTARAITTRSGQTLKSWWYNCP	1404
RESULT 6			
ID	AAU32262	AAU32262 standard; protein; 1415 AA.	
XX	AAU32262;		
AC	AAU32262;		
DT	18-DEC-2001	(first entry)	
DE	Novel human secreted protein #2753.		
KW	Human; vaccination; gene therapy; nutritional supplement;		
KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
OS	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
XX	Homo sapiens.		
EN	WO200179449-A2.		
XX	25-OCT-2001.		
PF	16-APR-2001; 2001WO-US008656.		
XX			

PR	18-APR-2000; 2000US-00552929.	
PR	26-JAN-2001; 2001US-00770160.	
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT;	
XX	WPI; 2001-611725/70.	
DR		
XX		
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic	
PT	vaccination, testing and therapy.	
XX		
PS	Claim 20; Page 573; 765pp; English.	
XX		
CC	The invention relates to novel human secreted polypeptides. The	
CC	polypeptides and antibodies to the polypeptides are useful for	
CC	determining the presence of or predisposition to a disease associated	
CC	with altered levels of polypeptide. The polypeptides are also useful for	
CC	identifying agents (agonists and antagonists) that bind to them. Cells	
CC	expressing the proteins are useful for identifying a therapeutic agent	
CC	for use in treatment of a pathology related to aberrant expression or	
CC	physiological interactions of the polypeptide. Vectors comprising the	
CC	nucleic acids encoding the polypeptides and cells genetically engineered	
CC	to express them are also useful for producing the proteins. The proteins	
CC	are useful in genetic vaccination, testing and therapy, and can be used	
CC	as nutritional supplements. They may be used to increase stem cell	
CC	proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon	
CC	and/or nerve tissue growth or regeneration; immune suppression and/or	
CC	stimulation; as anti-inflammatory agents; and in treatment of leukaemias.	
CC	AAU29510-AAU33304 represent the amino acid sequences of novel human	
CC	secreted proteins of the invention	
XX		
SQ	Sequence 1415 AA;	

Query Match	99.7%;	Score	7041.7;	DB	4;	Length	1415;
Best Local Similarity	93.3%;	Pred. No.	9.9e-171;				
Matches	1310;	Conservative	1;	Mismatches	0;	Indels	93;
Gaps	1;						
QY	1	MAWKTLPIVLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF	60				
Db	12	MAWKTLPIVLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF	71				
QY	61	KRVCTAELSCGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAE-----	106				
Db	72	KRVCTAELSCGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVINPTSPPESSKKAP	131				
QY	107	-----	106				
Db	132	PPSGASQIKSTTKRSPKPPNKKTKKVIIESEITEHVSVENQESSSSSSSSSSTIW	191				
QY	107	-----VKDNKNRTKKKPTKPPVVDAGSLDNGDFKVTTPDTST	147				
Db	192	KIKSSKNSAANRELQKKLVKDNKNRTKKKPTKPPVVDAGSLDNGDFKVTTPDTST	251				
QY	148	TOHNKVSTSPKITTAKPINRPSLPNNSDTSKETSLTVNKETTIVETKTTTTNKQTSIDG	207				
Db	252	TOHNKVSTSPKITTAKPINRPSLPNNSDTSKETSLTVNKETTIVETKTTTTNKQTSIDG	311				
QY	208	KEKTTSAKETQSTSEKTSKADLAPTSKVLAKEPTPKAETTTKGPALTPKPEPTTPPKGPAS	267				
Db	312	KEKTTSAKETQSTSEKTSKADLAPTSKVLAKEPTPKAETTTKGPALTPKPEPTTPPKGPAS	371				
QY	268	TTPKGPTPTTIKSAPTTPKPEAPTTTKSAPTTTPKPEAPTTTKPEAPTTPKPEAPTTTKPE	327				
Db	372	TTPKGPTPTTIKSAPTTPKPEAPTTTKSAPTTTPKPEAPTTTKPEAPTTPKPEAPTTTKPE	431				
QY	328	APTITTKSAPTTPKPEAPTTTPKAPTTTPKPEAPTTTPKPEPTTPPKGPAPTTKGPAPTTPK	387				
Db	432	APTITTKSAPTTPKPEAPTTTPKAPTTTPKPEAPTTTPKPEPTTPPKGPAPTTKGPAPTTPK	491				
QY	388	EPAPTAPKKPAPTTPKPEAPTTTPKPEAPTTTPKPEPTTPPKGPAPTTTKSAPTTTKGPAPT	447				
Db	492	EPAPTAPKKPAPTTPKPEAPTTTPKGPAPTTTKPEPTTPPKGPAPTTTKSAPTTTKGPAPT	551				

QY	448	TTKSAPTTTPKSPSPPTTTKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKP	507
Db	552	TTKSAPTTTPKSPSPPTTTKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKP	611
QY	508	APTAPKEPAPTTTPKATATTPKLTTPTPKEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT	567
Db	612	APTAPKEPAPTTTPKATATTPKLTTPTPKEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT	671
QY	568	PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAAPTTPKGTAPTTLKEP	627
Db	672	PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAAPTTPKGTAPTTLKEP	731
QY	628	APTTPKAPKAPKELAPTTTKEPTSTTSKDPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKG	687
Db	732	APTTPKAPKAPKELAPTTTKEPTSTTSKDPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKG	791
QY	688	TAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPTSTTSKDPAPTTTPKETAAPTTPKEPAPTTTPK	747
Db	792	TAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPTSTTSKDPAPTTTPKETAAPTTPKEPAPTTTPK	851
QY	748	KPAPTTTPPTTPPTTSEVSTPTTKEPTTTHKSPDESTPELSAETTPKALENSPKPSPGVT	807
Db	852	KPAPTTTPPTTPPTTSEVSTPTTKEPTTTHKSPDESTPELSAETTPKALENSPKPSPGVT	911
QY	808	TKTPAATKPEMTTAKDKTTERDLRTPPETTTAAAPKMTKETAATTTKTESKITATTQV	867
Db	912	TKTPAATKPEMTTAKDKTTERDLRTPPETTTAAAPKMTKETAATTTKTESKITATTQV	971
QY	868	TSSTTQDTPPFKITTLTKTLAPKVTTKTITTTTTEIMNKPEBTAKPKDRATNSKATTPK	927
Db	972	TSSTTQDTPPFKITTLTKTLAPKVTTKTITTTTTEIMNKPEBTAKPKDRATNSKATTPK	1031
QY	928	POKPTKAPKPTSTKPKTWPVRKPTTPPKMTSTWPELNPTTSRIAEAMLTQTTREN	987
Db	1032	POKPTKAPKPTSTKPKTWPVRKPTTPPKMTSTWPELNPTTSRIAEAMLTQTTREN	1091
QY	988	QTPNSKLVEVNPKESDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS	1047
Db	1092	QTPNSKLVEVNPKESDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS	1151
QY	1048	DETNIENGKPEVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPIPDVFT	1107
Db	1152	DETNIENGKPEVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPIPDVFT	1211
QY	1108	RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIIVAAALSTAKYXNWPBSVY	1167
Db	1212	RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIIVAAALSTAKYXNWPBSVY	1271
QY	1168	FFKRGSGIQOYIYKQBPVQKCPGRRPALNYPVYGEVMTQVRRRPERAIGPSQTHIRIQY	1227
Db	1272	FFKRGSGIQOYIYKQBPVQKCPGRRPALNYPVYGEVMTQVRRRPERAIGPSQTHIRIQY	1331
QY	1228	SPARLAYODKGVHLNEVKVLSILRGLPNVVTSAISLNPKNRPGDYDYAFSKOQYVNDV	1287
Db	1332	SPARLAYODKGVHLNEVKVLSILRGLPNVVTSAISLNPKNRPGDYDYAFSKOQYVNDV	1391
QY	1288	PSRTARAITTRSGQTLKSVWYNCP	1311
Db	1392	PSRTARAITTRSGQTLKSVWYNCP	1415
RESULT 7			
ID	ADK65839	standard; protein; 1404 AA.	
XX	AC	ADK65839;	
XX	DT	06-MAY-2004 (first entry)	
DE	XX	Angiogenesis-differentially expressed protein #53.	
XX	XX	cytostatic; cardiant; vasotropic; antiarteriosclerotic;	
KW	XX		

angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
gene expression; cancer; coronary artery disease; myocardial ischemia;
coronary arteriosclerosis; forensic medicine.
Homo sapiens.
WO2003066831-A2.
14-AUG-2003.
07-FEB-2003; 2003WO-US003848.
07-FEB-2002; 2002US-00067482.
10-JUN-2002; 2002US-00164595.
16-AUG-2002; 2002US-0403649P.
03-JAN-2003; 2003US-0437746P.
(ORIG-) ORIGENE TECHNOLOGIES INC.
Sun Z, Li X, Kovacs KF, Fan W, Jay G;
WPI; 2003-731502/59.
Determining the angiogenic index of a tissue or cell sample using
expression levels of differentially expressed genes, useful for
diagnosing or treating cancer, coronary artery disease, myocardial
ischemia and/or arteriosclerosis.
Disclosure; SEQ ID NO 78; 296pp; English.
The invention relates to a method of determining the angiogenic index of
a tissue or cell sample comprising assessing, in a sample, the expression
levels of one or more differentially-expressed gene from any of 34 DNA
sequences, given in the specification, where the levels are indicative of
the angiogenic index. The methods and compositions of the present
invention are useful for diagnosing, preventing and/or treating cancer,
coronary artery disease, myocardial ischemia or coronary
arteriosclerosis. They can also be used in research, drug discovery and
forensic medicine involving angiogenesis. This sequence corresponds to
one of the differentially expressed proteins of the invention.

Query Match 99.5%; Score 7029.7; DB 7; Length 1404;
Best Local Similarity 93.2%; Pred. No. 2e-170;
Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFERGRCDCDAQCKKYDKCCPDYEFCAE----- 106
Db 61 KRVTAEISCKGRCFESFERGRCDCDAQCKKYDKCCPDYEFCAEVHNTPSPSSKAP 120
QY 107 ----- 106
Db 121 PPGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIR 180
QY 107 -----VKDNKNPTKKKPKPPVVDVDEAGSLNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNKNTKKKPKPPVVDVDEAGSLNGDFKVTTPDTST 240
QY 148 TQHNKVTSPKITTAKPINRPSLPNNSDTSKETSILTVNKETTIVETKTTTNKQTSIDG 207
Db 241 TQHNKVTSPKITTAKPINRPSLPNNSDTSKETSILTVNKETTIVETKTTTNKQTSIDG 300
QY 208 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKSPAS 267
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKSPAS 360
QY 268 TTPKEPTTPKSPAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKPE 327
Db |||||


```
QY 1 MAWKTLPIYLLILLLSVFIQVSSODLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLILLLSVFIQVSSQ-----25
QY 61 KRVCTAELSCGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVKDNKNRTKKKPT 120
Db 26 -----EUSCKGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVKDNKNRTKKKPT 79
QY 121 PKPPVVVDAGSLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNPSDTSKE 180
Db 80 PKPPVVVDAGSLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNPSDTSKE 139
QY 181 TSLTVNKETTVTKETTTNKOTSDGKEKTTSAKETQSIEKTSADIAPTSKVLAKPTP 240
Db 140 TSLTVNKETTVTKETTTNKOTSDGKEKTTSAKETQSIEKTSADIAPTSKVLAKPTP 199
QY 241 KAEITTKGPAITPKPEPTTTPKEPASTTPKEPTTTIKSAPTTPKPEAPTTTKSAPITP 300
Db 200 KAEITTKGPAITPKPEPTTTPKEPASTTPKEPTTTIKSAPTTPKPEAPTTTKSAPITP 259
QY 301 KEPAITTTKEPAITPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKPEAPTTKEPAP 360
Db 260 KEPAITTTKEPAITPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKPEAPTTKEPAP 319
QY 361 TTPKGPPTTPKEPAITPKPEAPTTKEPAPTAPKPAITPKPEAPTTPKPEAPTTTKE 420
Db 320 TTPKGPPTTPKEPAITPKPEAPTTKEPAPTAPKPAITPKPEAPTTPKPEAPTTTKE 379
QY 421 PSPTTPKEPAITTKSAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTKEPAPITP 480
Db 380 PSPTTPKEPAITTKSAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTKEPAPITP 439
QY 481 KKPAITPKPEAPTTTKPEAPTTTKKPAITAPKPEAPTTPKETAPTTPKKLPTTPEKLA 540
Db 440 KKPAITPKPEAPTTTKPEAPTTTKKPAITAPKPEAPTTPKETAPTTPKKLPTTPEKLA 499
QY 541 PTTPEKPAITPEELAPTPEPTTTPPEAPTTTPKAAAPNTPKPEAPTTPKPEAPITP 600
Db 500 PTTPEKPAITPEELAPTPEPTTTPPEAPTTTPKAAAPNTPKPEAPTTPKPEAPITP 559
QY 601 KEPAITPKETAPTTPKGTAPTTLKEPAITPKKPAKELAPTTPKEPTSTTSDDKPAIT 660
Db 560 KEPAITPKETAPTTPKGTAPTTLKEPAITPKKPAKELAPTTPKEPTSTTSDDKPAIT 619
QY 661 PKGTAPTTPKEPAITPKGPAPTTLKEPAITPKKPAKELAPTTPKGTST 720
Db 620 PKGTAPTTPKEPAITPKGPAPTTLKEPAITPKKPAKELAPTTPKGTST 679
QY 721 TSDKPAITPKETAPTTPKEPAITPKKPAITPKETPTTSEVSTPTTKEPTTIHKSP 780
Db 680 TSDKPAITPKETAPTTPKEPAITPKKPAITPKETPTTSEVSTPTTKEPTTIHKSP 739
QY 781 DESTPELSAEPTEKALENSKEPGVPTTKPAATKPEMTTAKDKITERDLRTPETTTA 840
Db 740 DESTPELSAEPTEKALENSKEPGVPTTKPAATKPEMTTAKDKITERDLRTPETTTA 799
QY 841 APRMTKETATTTBKTTESKITATTTQVSTSTTQDTPPKITLTKTTTLAPKVTITTKIT 900
Db 800 APRMTKETATTTBKTTESKITATTTQVSTSTTQDTPPKITLTKTTTLAPKVTITTKIT 859
QY 901 TTEIMNKPEETAKPKORATNSKATTPKOKETKAPKPTSTKPKTMPVRVKPTTTPR 960
Db 860 TTEIMNKPEETAKPKORATNSKATTPKOKETKAPKPTSTKPKTMPVRVKPTTTPR 919
QY 961 KMTSTWPELNPSTRIAEAMLQTTTRPNQTPNSKLIVEVNPXSEDAGGAEGETPHMLPHV 1020
Db 920 KMTSTWPELNPSTRIAEAMLQTTTRPNQTPNSKLIVEVNPXSEDAGGAEGETPHMLPHV 979
QY 1021 FMEVTPDMOYLPRVFNQGIINPMLSDNETNICNGKVDGLTTLRNGTLVAFRGHYFWM 1080
Db 980 FMEVTPDMOYLPRVFNQGIINPMLSDNETNICNGKVDGLTTLRNGTLVAFRGHYFWM 1039
```

1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYMRFTNDIKDAGYKPIF 1140
1040 SPFSPSPARRITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYMRFTNDIKDAGYKPIF 1099
1141 KGFGGLTQIIVAAALSTAKYKNWPSVYFFKGGSTQOVIYKQEPVOKCGRRPALNYPV 1200
1100 KGFGGLTQIIVAAALSTAKYKNWPSVYFFKGGSTQOVIYKQEPVOKCGRRPALNYPV 1159
1201 GEMTOVRRRRFERAIGPSQTHIRIOYSPARLAYODKGVLNHNEVKVSLWRGLPNVISA 1260
1160 GETTOVRRRRFERAIGPSQTHIRIOYSPARLAYODKGVLNHNEVKVSLWRGLPNVISA 1219
1261 ISLPNIRKPDGYDYAFSKDQYVNDVPSRTARAITTRSGOTLSKVWYNCP 1311
1220 ISLPNIRKPDGYDYAFSKDQYVNDVPSRTARAITTRSGOTLSKVWYNCP 1270

RESULT 9
ADK65819
ID ADK65819 standard; protein; 1320 AA.
XX
AC ADK65819;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0316.
XX
KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
PN WC2003066831-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WC-US003848.
XX
PR 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX
N-PSDB; ADK65818.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 58; 296pp; English.
XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present
CC invention are useful for diagnosing, preventing and/or treating cancer,
CC coronary artery disease, myocardial ischemia or coronary
CC arteriosclerosis. They can also be used in research, drug discovery and
CC forensic medicine involving angiogenesis. This sequence corresponds to
CC one of the differentially expressed proteins of the invention.
XX
SQ Sequence 1320 AA;

Query Match 95.8%; Score 6769.9; DB 7; Length 1320;

Best Local Similarity 93.1%; Pred. No. 7.3e-164;
Matches 1267; Conservative 0; Mismatches 3; Indels 91; Gaps 2;
QY 1 MAWTLPIYLLLLLSVFIQOVSSQDLSSCAGRCGEGYSDRATCNDYNCOHYMECCPDF 60
Db 1 MAWTLPIYLLLLLSVFIQOVSSQ----- 25
QY 61 KRVCIAELSKGRCFESFERGECDDAQCKYDKCCPDYEFCA----- 105
Db 26 -----ELSKGRCFESFERGECDDAQCKYDKCCPDYEFCAEYVHNTPSPSSKKAP 79
QY 106 -----EVKDKNRTKKKPTPKPPVDEAG 130
Db 80 PPSGASQTIKSTTKRSPKPNKKTKKVIKIESEITEVKDKNRTKKKPTPKPPVDEAG 139
QY 131 SGLDNGDFKVTTPDTSSTQHNKYSTSPKITTAKPINRPSISLPNSDTSKETSLSLVNKETT 190
Db 140 SGLDNGDFKVTTPDTSSTQHNKYSTSPKITTAKPINRPSISLPNSDTSKETSLSLVNKETT 199
QY 191 VETKETTNTKQSTDKGKXTTSKAKETQSIKTSKADLAPTSKVLAKPTKAEATTTKGPA 250
Db 200 VETKETTNTKQSTDKGKXTTSKAKETQSIKTSKADLAPTSKVLAKPTKAEATTTKGPA 259
QY 251 LTTPEKERTPTPKEPASTTTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 310
Db 260 LTTPEKERTPTPKEPASTTTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 319
QY 311 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 370
Db 320 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 379
QY 371 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 430
Db 380 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 439
QY 431 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 490
Db 440 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 499
QY 491 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 550
Db 500 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 559
QY 551 TPEELAPTTPEEPTPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 610
Db 560 TPEELAPTTPEEPTPTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 619
QY 611 TAPTTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 670
Db 620 TAPTTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 679
QY 671 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 730
Db 680 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 739
QY 731 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 790
Db 740 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 799
QY 791 PTPKALENSPKEPGVPTTKTTPAATKPEMTTAKDKITRDLTTPETTAAAPKMTKETAT 850
Db 800 PTPKALENSPKEPGVPTTKTTPAATKPEMTTAKDKITRDLTTPETTAAAPKMTKETAT 859
QY 851 TTEKTTESKITATTTQVSTTTQDTTPPKITTLTKTTTLAPKVTTTKKITTITTEIMNKPEE 910
Db 860 TTEKTTESKITATTTQVSTTTQDTTPPKITTLTKTTTLAPKVTTTKKITTITTEIMNKPEE 919
QY 911 TAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTMVRVKPTTTPRKMSTSMPELN 970
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTMVRVKPTTTPRKMSTSMPELN 979
QY 971 PTSRIAEMLQTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1030

Db 980 PTSRIAEMLQTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1031 YLPRVPMQGIINPMLSDETNICNGKPVVDGLTTILRNGLTVAFRGHYFWMLSPFSPSPAR 1090
Db 1040 YLPRVPMQGIINPMLSDETNICNGKPVVDGLTTILRNGLTVAFRGHYFWMLSPFSPSPAR 1099
QY 1091 RITEVWGIPIPIDITVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1150
Db 1100 RITEVWGIPIPIDITVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1159
QY 1151 VAALSTAKYKNWSPESVYFFKRGSGIQYIYKQBPVQKCPGRRPALNYPVYVGTQVRRR 1210
Db 1160 VAALSTAKYKNWSPESVYFFKRGSGIQYIYKQBPVQKCPGRRPALNYPVYVGTQVRRR 1219
QY 1211 FERAIGSQHTTIRIQYSPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPD 1270
Db 1220 FERAIGSQHTTIRIQYSPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPD 1279
QY 1271 GYDYAFSKDQYVYNDVPSTARAITTRSQTLTSKWYNCP 1311
Db 1280 GYDYAFSKDQYVYNDVPSTARAITTRSQTLTSKWYNCP 1320
RESULT 10
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX AC AAM24322;
XX DT 12-OCT-2001 (first entry)
XX DE Human EST encoded protein SEQ ID NO: 1847.
XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
XX KW gene therapy; nutrition.
XX OS Homo sapiens.
XX PN WO200154477-A2.
XX PN 02-AUG-2001.
XX PD 25-JAN-2001; 2001WO-US002687.
XX PF 25-JAN-2000; 2000US-00491404.
XX PR 17-JUL-2000; 2000US-00617746.
XX PR 03-AUG-2000; 2000US-00631451.
XX PR 15-SEP-2000; 2000US-00663870.
XX (HYSE-) HYSEQ INC.
XX PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Wehrman T;
XX WPI; 2001-476164/51.
XX DR N-PSDB; AAH98981.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
antibodies and research use.
PS Claim 20; Page 1198-1201; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
proteins from a variety of organisms, including human, dog, cat, horse,
cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
urchin and tomato. These were derived from expressed sequence tags (ESTs)
from the organism of interest. They can be used in diagnostics,
forensics, gene mapping, identification of mutations, to assess
biodiversity and for nutritional purposes. The present sequence is a
protein of the invention
XX

SQ	Sequence 1299 AA;	
	Query Match	91.6%; Score 6468.7; DB 4; Length 1299;
	Best Local Similarity	92.8%; Pred. No. 3.3e-156;
	Matches 1202; Conservative	0; Mismatches 0; Indels 93; Gaps 1;
QY	1 MAWKTLPIYLLLLLVFVJQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60
DB	1 MAWKTLPIYLLLLLVFVJQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF	60
QY	61 KEVCTAEUSCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAE-----	106
DB	61 KRVCFAELUSCKGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVHNPTSPPSKKAP	120
QY	107 -----	106
DB	121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEBITEHSVSENQESSSSSSSSSTIW	180
QY	107 -----VKONKNRTYKKKTPKPPVVDGAGSLDNGDFKVTTPDTST	147
DB	181 KIKSSKNSAANRELOKQLVKONKNRTYKKKTPKPPVVDGAGSLDNGDFKVTTPDTST	240
QY	148 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKEITTTNKQSTDG	207
DB	241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKEITTTNKQSTDG	300
QY	208 KBKTTSAKETQSIEKTSAXDLAPTSKVLAKPTPKAETTTKGPALTTPKGPTTTPKEPAS	267
DB	301 KEKTTSAKETQSIEKTSAXDLAPTSKVLAKPTPKAETTTKGPALTTPKGPTTTPKEPAS	360
QY	268 TTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	327
DB	361 TTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	420
QY	328 APPTTKSAPTTPKEPAPTTPKAPPTPKAPPTTPKEPTTTPKGPAPTTKBPAPTTPK	387
DB	421 APPTTKSAPTTPKEPAPTTPKAPPTPKAPPTTPKEPTTTPKGPAPTTKBPAPTTPK	480
QY	388 EPAPTAPKPAPTTPKEPAPTTPKAPPTTKSPSTTPKEPAPTTTKSAPTTTKEPAPT	447
DB	481 EPAPTAPKPAPTTPKEPAPTTPKAPPTTKSPSTTPKEPAPTTTKSAPTTTKEPAPT	540
QY	448 TTKSAPTTPKEPPTTTKEPAPTTPKAPPTTPKAPPTTPKEPAPTTTPKEPAPTTTKP	507
DB	541 TTKSAPTTPKEPPTTTKEPAPTTPKAPPTTPKAPPTTPKEPAPTTTPKEPAPTTTKP	600
QY	508 APAPKPAPTTPKEPAPTTPKKLTPTTPEKLAFTTPKEPAPTTPEELAPTTPEEPTPTT	567
DB	601 APAPKPAPTTPKEPAPTTPKKLTPTTPEKLAFTTPKEPAPTTPEELAPTTPEEPTPTT	660
QY	568 PEPAPPTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTTLKEP	627
DB	661 PEPAPPTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTPKGTAPTTLKEP	720
QY	628 APPTPKAPKELAPTTTKBPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG	687
DB	721 APPTPKAPKELAPTTTKBPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG	780
QY	688 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTTPKEPAPTTPK	747
DB	781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTTPKEPAPTTPK	840
QY	748 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSDESTPELSAETPKALENSKPEGVPT	807
DB	841 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSDESTPELSAETPKALENSKPEGVPT	900
QY	808 TKTPAATKPEMTTAKDKXTERDLRTTPETTTAAPKMTKATATTEKTTESKIATTTQV	867
DB	901 TKTPAATKPEMTTAKDKXTERDLRTTPETTTAAPKMTKATATTEKTTESKIATTTQV	960
QY	868 TSTTTQDTPPKITTLKTTTLAPKVTTKXITTTTEIMNKPEEETAKPKDRATNSKATTPK	927
DB	961 TSTTTQDTPPKITTLKTTTLAPKVTTKXITTTTEIMNKPEEETAKPKDRATNSKATTPK	1020

QY	928 POKPTKAPKPTSTKKPTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRN	987
DB	1021 POKPTKAPKPTSTKKPTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRN	1080
QY	988 QTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1047
DB	1081 QTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1140
QY	1048 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIPSPIDTFT	1107
DB	1141 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIPSPIDTFT	1200
QY	1108 RCNCGKTPFFPKDSQWRFNTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESVY	1167
DB	1201 RCNCGKTPFFPKDSQWRFNTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESVY	1260
QY	1168 FFKRGSGIQYIYKQEPVQKCPGRPALNYPVYGE	1202
DB	1261 FFKRGSGIQYIYKQEPVQKCPGRPALNYPVYGE	1295
RESULT 11		
AAB29778		
ID	AAB29778 standard; protein; 902 AA.	
XX		
AC	AAB29778;	
XX		
DT	28-FEB-2001 (first entry)	
XX		
DE	Human MSF-derived tribonection.	
XX		
KW	Human tribonection; MSF; megakaryocyte stimulating factor;	
KW	alternative splicing; joint boundary lubricant; O-linked oligosaccharide;	
KW	osteoarthritis; tribosupplementation; tissue adhesion inhibition;	
KW	friction coefficient reduction; gene therapy; antiarthritic; osteopathic.	
OS	Homo sapiens.	
XX		
PN	WO200064930-A2.	
XX		
PD	02-NOV-2000.	
XX		
PF	24-APR-2000; 2000WO-US010953.	
XX		
PR	23-APR-1999; 99US-00298970.	
XX		
PA	(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.	
XX		
PI	Jay GD;	
XX		
DR	WPI; 2001-024673/03.	
XX		
PT	Novel tribonection polypeptide useful as lubricant for treating	
XX	osteoarthritis, comprises O-linked lubricating moiety.	
PS	Disclosure; Fig 1; 47pp; English.	
XX		
CC	The invention relates to a human tribonection which is a product of	
CC	alternative splicing of the human MSF (megakaryocyte stimulating factor)	
CC	gene. The tribonection has at least one O-linked oligosaccharide	
CC	lubricating moiety and has a polypeptide sequence comprising 1-76 repeats	
CC	of a motif having at least 50% identity to the sequence KEFAPTT	
CC	(AAB29774). The invention also relates to a nucleic acid encoding a human	
CC	MSF-derived tribonection; a biocompatible composition comprising a human	
CC	tribonection for inhibiting tissue adhesion formation; and a method of	
CC	diagnosing osteoarthritis or a predisposition to osteoarthritis by	
CC	measuring the amount of MSF or its fragment in a biological sample of a	
CC	mammal, wherein an increased amount of MSF compared to a control	
CC	indicates the presence of or predisposition to developing osteoarthritis.	
CC	The tribonection and DNA encoding it are useful in the treatment of	
CC	osteoarthritis, where they may be used for lubricating mammalian joints,	
CC	such as articulating joints of humans, dogs or horses. The tribonection,	

QY 624 LKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPT 683
DB 121 LKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPT 180
QY 684 TPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPT 743
DB 181 TPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPT 240
QY 744 TTPKKPAPTTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEP 803
DB 241 TTPKKPAPTTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEP 300
QY 804 GVPTTKTAATKPEMTTAKDKTTTTERDLRTTTPETTTAAPKMTKETATTTTEKTESKITAT 863
DB 301 GVPTTKTAATKPEMTTAKDKTTTTERDLRTTTPETTTAAPKMTKETATTTTEKTESKITAT 360
QY 864 TTQVTSITTTQDTPPKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKA 923
DB 361 TTQVTSITTTQDTPPKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKA 420
QY 924 TTPKPKQPTKAPKKPTSTTKPKTMPRVRKPTTTPRKMSTMPPELNPTSRJAEMLOTT 983
DB 421 TTPKPKQPTKAPKKPTSTTKPKTMPRVRKPTTTPRKMSTMPPELNPTSRJAEMLOTT 480
QY 984 TRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 1043
DB 481 TRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 540
QY 1044 PMLSDETNI:CN 1054
DB 541 PMLSDETNI:CN 551
RESULT 13
ID ABUS3252 standard; protein; 546 AA.
XX AC ABUS3252;
DT 14-APR-2003 (first entry)
XX DE Human testes-derived DKFphtes3_4o19 homologue #1.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WC200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX PI WPI; 2001-327840/34.
XX DR Nucleic acids having the sequences of clones isolated from libraries of
XX PT different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 892; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and

CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 546 AA;
Query Match 41.3%; Score 2920; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.8e-66;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 395 KKPAPTTTKEPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPT 454
DB 1 KKPAPTTTPKEPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPT 60
QY 455 TPKEPSPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 514
DB 61 TPKEPSPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 120
QY 515 PAPTTPKETAPTTPKKLTPTTPEKLAPTTTPKEPAPTTPEELAPTTPPEEPTTTPPEEAPT 574
DB 121 PAPTTPKETAPTTPKKLTPTTPEKLAPTTTPKEPAPTTPEELAPTTPPEEPTTTPPEEAPT 180
QY 575 TPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 634
DB 181 TPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 240
QY 635 PAKKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPTTPKGTAPTTLK 694
DB 241 PAKKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPTTPKGTAPTTLK 300
QY 695 EPAPTTTPKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPTTP 754
DB 301 EPAPTTTPKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKAPATTTPKEPAPTTP 360
QY 755 ETTPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAETPKALENSPKEPGVPTTKPAAT 814
DB 361 ETTPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAETPKALENSPKEPGVPTTKPAAT 420
QY 815 KPEMTTAKDKTTTERDLRTTTPETTTAAPKMTKETATTTTEKTESKITATTTQVTSITTD 874
DB 421 KPEMTTAKDKTTTERDLRTTTPETTTAAPKMTKETATTTTEKTESKITATTTQVTSITTD 480
QY 875 TTPFKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 934
DB 481 TTPFKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540
QY 935 PKKPTS 940
DB 541 PKKPTS 546
RESULT 14
AAO18834
ID AAO18834 standard; protein; 538 AA.
XX AC AAO18834;
XX DT 29-OCT-2002 (first entry)
XX DE 3' cartilage superficial zone protein coding sequence encoded protein.
XX KW SZP; superficial zone protein; cartilage; lubrication; human;
XX KW degenerative joint condition; arthritis; osteoporosis; trauma; CACF;
XX KW chondroicin sulphate substitution consensus; antiarthritic;
XX KW antirheumatic; osteopathic; antilgout; antiinflammatory; dermatological;
XX KW immunosuppressive.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT Misc-difference 114 /note= "encoded by ACTACT"

Wed Oct 13 12:38:28 2004

Db	241	TPKEPAPTTTKEPTTTPKEPAPTTTKEPAPTTAPKKPAPTTKEPAPTTTPKEP	300
Qy	414	APTITKESPTTPKEPAPTTTTSAPTITKESAPTITKESPTTITKEPAPTTPK	473
Db	301	APTITKESPTTPKEPAPTTTTSAPTITKESAPTITKESPTTITKEPAPTTPK	360
Qy	474	EPAPTTPKKAPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKKTLP	533
Db	361	EPAPTTPKKAPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKKTLP	420
Qy	534	TTPKLAAPTTPKAPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK	593
Db	421	TTPKLAAPTTPKAPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK	480
Qy	594	EPAPTTKEPAPTTTPKETAPTTTPKETAPTTLKE	626
Db	481	EPAPTTKEPAPTTTPKETAPTTTPKETAPTTLKE	513

Search completed: October 13, 2004, 11:37:31
Job time : 114.696 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.7891 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-H

Perfect score: 7064

Sequence: 1 MAWKTLPIVILLLSVFVIQ.....ARAITRSGQTLSKWNCP 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5A COMB.pdp.*
- 2: /cgn2_6/prodata/1/iaa/5B COMB.pdp.*
- 3: /cgn2_6/prodata/1/iaa/6A COMB.pdp.*
- 4: /cgn2_6/prodata/1/iaa/6B COMB.pdp.*
- 5: /cgn2_6/prodata/1/iaa/PCUS COMB.pdp.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7064	100.0	1311	4	US-07-757-022B-42
2	7049.7	99.8	1354	4	US-07-757-022B-48
3	7049	99.8	1361	4	US-07-757-022B-40
4	7044.7	99.7	1404	4	US-07-757-022B-2
5	7044.7	99.7	1404	4	US-07-757-022B-62
6	7044.7	99.7	1404	4	US-09-298-970A-1
7	7029.7	99.5	1404	4	US-10-164-595-78
8	6811.3	96.4	1314	4	US-07-757-022B-50
9	6799.9	96.3	1270	4	US-07-757-022B-44
10	6785.6	96.1	1313	4	US-07-757-022B-142
11	6784.9	96.0	1320	4	US-07-757-022B-46
12	6784.9	96.0	1320	4	US-07-757-022B-60
13	6780.6	96.0	1363	4	US-07-757-022B-52
14	6769.9	95.8	1320	4	US-10-164-595-58
15	5603.7	79.3	1140	4	US-07-757-022B-104
16	5500.7	77.9	1038	4	US-07-757-022B-74
17	5472.6	77.5	1022	4	US-07-757-022B-84
18	5344.6	75.7	1049	4	US-07-757-022B-58
19	5011	70.9	941	4	US-07-757-022B-14
20	2057.8	29.1	463	4	US-07-757-022B-54
21	1806.8	25.6	423	4	US-07-757-022B-68
22	1793.1	24.4	422	4	US-07-757-022B-66
23	1723.1	24.4	372	4	US-07-757-022B-64
24	1398.2	19.8	5179	4	US-09-538-092-1258
25	1219.6	17.3	237	4	US-07-757-022B-72
26	1139.7	16.1	8991	4	US-08-714-741-32
27	1108.5	15.7	296	4	US-07-757-022B-70

28 884 12.5 157 4 US-07-757-022B-102
29 884 12.5 157 4 US-07-757-022B-114
30 869 12.3 207 4 US-07-757-022B-116
31 869 12.3 207 4 US-07-757-022B-136
32 849.6 12.0 3118 3 US-09-579-181-1
33 842.3 11.9 2972 3 US-09-579-181-2
34 824.7 11.7 3256 4 US-09-919-172-98
35 824.7 11.7 3256 4 US-09-976-594-22
36 824.7 11.7 3256 4 US-09-919-039-21
37 814.8 11.5 1837 3 US-08-928-361B-5
38 814.8 11.5 1837 4 US-09-588-995A-5
39 808.9 11.5 2142 4 US-09-538-092-1142
40 808.4 11.4 4019 4 US-09-854-133-425
41 794.6 11.2 1721 3 US-08-700-651-5
42 794.6 11.2 1721 3 US-08-928-361B-6
43 794.6 11.2 1721 4 US-09-588-995A-6
44 771 10.9 231 4 US-07-757-022B-30
45 763 10.8 132 4 US-07-757-022B-140

ALIGNMENTS

RESULT 1
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID

; TOPOLOGY: linear	
; MOLECULE TYPE: protein	
US-07-757-022B-42	
Query Match 100.0%; Score 7064; DB 4; Length 1311;	
Best Local Similarity 100.0%; Pred. No. 4.5e-193;	
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
DB 1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
QY 61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVKDKNKKRKKPT 120
DB 61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVKDKNKKRKKPT 120
QY 121	KRPVVVDEAGSLGNDGKVTTPDTSTOHNVKSTSPKLTAKPINRPSLPNSDTSKE 180
DB 121	KRPVVVDEAGSLGNDGKVTTPDTSTOHNVKSTSPKLTAKPINRPSLPNSDTSKE 180
QY 181	TSLTVNKEITTVETKETTITNKTSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 240
DB 181	TSLTVNKEITTVETKETTITNKTSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 240
QY 241	KAETTTKGALTPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300
DB 241	KAETTTKGALTPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300
QY 301	KEPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPAP 360
DB 301	KEPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPAP 360
QY 361	TTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTKE 420
DB 361	TTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTKE 420
QY 421	PSGTTPEKPAATTTKAPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 480
DB 421	PSGTTPEKPAATTTKAPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 480
QY 481	KKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEKLA 540
DB 481	KKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEKLA 540
QY 541	PTTPEKPAATTPPEELAPTTPEEPTTPPEEPTTPPEEPTTPKAAAPNTPKGPAPTTKEPAPTTP 600
DB 541	PTTPEKPAATTPPEELAPTTPEEPTTPPEEPTTPPEEPTTPKAAAPNTPKGPAPTTKEPAPTTP 600
QY 601	KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTKEPTSTTSKDPAPT 660
DB 601	KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTKEPTSTTSKDPAPT 660
QY 661	PKGTAPTTPKEPATTTPKEPATTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKGPTST 720
DB 661	PKGTAPTTPKEPATTTPKEPATTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKGPTST 720
QY 721	TSDKPAATTPKETAPTTPKGPAPTTPKGPAPTTPETPPPTTSEVSTTTKEPTTIHKSP 780
DB 721	TSDKPAATTPKETAPTTPKGPAPTTPKGPAPTTPETPPPTTSEVSTTTKEPTTIHKSP 780
QY 781	DESTPELSAETTPKALSNSKEPGVPTTKPAATKPEMTTTAKDKTTERDLRTTPETTTA 840
DB 781	DESTPELSAETTPKALSNSKEPGVPTTKPAATKPEMTTTAKDKTTERDLRTTPETTTA 840
QY 841	APKMKETATTTETKTTESKITAATTTQVTSSTTTQDTPFKITLTKTTLAPKVTTKKTI 900
DB 841	APKMKETATTTETKTTESKITAATTTQVTSSTTTQDTPFKITLTKTTLAPKVTTKKTI 900
QY 901	TTEIMNKPEETAKPKORATNSKATTPPKOKTKAPKPTSTKKPKTPRVRKPKTTPTPR 960
DB 901	TTEIMNKPEETAKPKORATNSKATTPPKOKTKAPKPTSTKKPKTPRVRKPKTTPTPR 960
QY 961	KWTSTMPELNPTSIAEAMLQTTTRPNQTPNSKLVENPKSEDAGGABGETPHMLLRPHV 1020

```
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-48

Query Match      99.8%; Score 7049.7; DB 4; Length 1354;
Best Local Similarity 96.8%; Pred. No. 1.2e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQOVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQOVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
QY 61 KRVTAE LSCGRCFESFERGRCDCDAQCKYDKCCPDYSEFCAE-----106
DB 61 KRVTAE LSCGRCFESFERGRCDCDAQCKYDKCCPDYSEFCAEHSVSENOESSSS 120
QY 107 -----VKONKQRTKKKPTKPPVVDAGSLDNGD 137
DB 121 SSSSSSTIWKIKSSKNSAANRELQKLKVKONKQRTKKKPTKPPVVDAGSLDNGD 180
QY 138 FKVTTPDTSHTQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKEIT 197
DB 181 FKVTTPDTSHTQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKEIT 240
QY 198 TTNKQTS DGEKNTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPA LTTPKEP 257
DB 241 TTNKQTS DGEKNTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPA LTTPKEP 300
QY 258 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTPK 317
DB 301 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTPK 360
QY 318 EPAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSA 437
DB 421 TKEPAPTTTPKEPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480
QY 438 PTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 497
DB 481 PTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540
QY 498 EPAPTTTKKPAAPAPKEPAPTTPKETAPTTPKLPTTPEKLA PTTPEKAPTTPEELAP 557
DB 541 EPAPTTTKKPAAPAPKEPAPTTPKETAPTTPKLPTTPEKLA PTTPEKAPTTPEELAP 600
QY 558 TTPPEPTPTTPKEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 617
DB 601 TTPPEPTPTTPKEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 660
QY 618 GTAPTTTPKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 677
DB 661 GTAPTTTPKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 720
QY 678 KEPAPTTTPKGA PTTTPKEPAPTTTPKAPKELAPTTTPKGPSTTSDKAPTTTPKEPAPTTTPK 737
DB 721 KEPAPTTTPKGA PTTTPKEPAPTTTPKAPKELAPTTTPKGPSTTSDKAPTTTPKEPAPTTTPK 780
QY 738 PKEPAPTTTPKGA PTTTPKEPAPTTTPKAPKELAPTTTPKGPSTTSDKAPTTTPKEPAPTTTPK 797
DB 781 PKEPAPTTTPKGA PTTTPKEPAPTTTPKAPKELAPTTTPKGPSTTSDKAPTTTPKEPAPTTTPK 840
QY 798 NSPEKPGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKWTKEATTEKTE 857
DB 841 NSPEKPGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKWTKEATTEKTE 900
QY 858 SKITATTTQVTSITTTQDTPPKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDR 917
DB 901 SKITATTTQVTSITTTQDTPPKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDR 960

RESULT 3
US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (617) 876-1170	
TELEFAX: (617) 876-5851	
INFORMATION FOR SEQ ID NO: 40:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1361 amino acids	
TYPE: AMINO ACID	
TOPOLOGY: linear	
MOLECULE TYPE: protein	
US-07-757-022B-40	
Query Match	
Best Local Similarity 99.8%; Score 7049; DB 4; Length 1361;	
Matches 1311; Conservative 0; Mismatches 0; Indels 50; Gaps 1;	
QY	1 MAWKTLPIYLLLSLVFVYQVSSQDLSSCAGRCGEGYSRDATCNQYNCQHYMECCPDF 60
DB	1 MAWKTLPIYLLLSLVFVYQVSSQDLSSCAGRCGEGYSRDATCNQYNCQHYMECCPDF 60
QY	61 KRVTAEIACKGRCFESFGRGECDCDAQCKYDKCCPDYESFCA----- 105
DB	61 KRVTAEIACKGRCFESFGRGECDCDAQCKYDKCCPDYESFCA----- 105
QY	106 -----EVKDKKQRTKKKTKKVIKESBEITVKKDKKRTKKKTKKPPVVDEAG 130
DB	121 PPSGASQTIKSTTKRSPKPNKKTKKVIKESBEITVKKDKKRTKKKTKKPPVVDEAG 180
QY	131 SGLDNGDFKVTPTDSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKET 190
DB	181 SGLDNGDFKVTPTDSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKET 240
QY	191 VETKETTNTKQSTIDGKEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPA 250
DB	241 VETKETTNTKQSTIDGKEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPA 300
QY	251 LPTPKETPTPKEPASTTPKBPPTTIKSAPTPKEPAPTTTKSAPTPKEPAPTTTKE 310
DB	301 LPTPKETPTPKEPASTTPKBPPTTIKSAPTPKEPAPTTTKSAPTPKEPAPTTTKE 360
QY	311 PAPTTPKEPAPTTTKEPAPTTTSAPTPKEPAPTTPKAPPTPKAPPTPKAPPTTPT 370
DB	361 PAPTTPKEPAPTTTKEPAPTTTSAPTPKEPAPTTPKAPPTPKAPPTPKAPPTTPT 420
QY	371 PKEPAPTTKEPAPTTTKEPAPTTKAPKAPPTPKEPAPTTTKEPAPTTTKEPAPTTKEPA 430
DB	421 PKEPAPTTKEPAPTTTKEPAPTTKAPKAPPTPKEPAPTTTKEPAPTTTKEPAPTTKEPA 480
QY	431 PTTTKSAPTTTKEPAPTTTSAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 490
DB	481 PTTTKSAPTTTKEPAPTTTSAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
QY	491 PAPTTPKEPAPTTTKEPAPTTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 550
DB	541 PAPTTPKEPAPTTTKEPAPTTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600
QY	551 TPPELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 610
DB	601 TPPELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660
QY	611 TAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 670
DB	661 TAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720
QY	671 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 730
DB	721 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780
QY	731 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 790
DB	781 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 840
QY	791 PTPKALENSPKGPGVPTTKTAAKPEMTTAKDKTTERDLRTPETTTAAPKMTKETAT 850
DB	841 PTPKALENSPKGPGVPTTKTAAKPEMTTAKDKTTERDLRTPETTTAAPKMTKETAT 900

RESULT 4
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION DATA: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-2

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 1.8e-192;

Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYVLLLSVFIQVSSQDLSSCAGRCGEGYRDATCNCYDNCQHYWECPCDF 60
 DB 1 MAWKTLPIYVLLLSVFIQVSSQDLSSCAGRCGEGYRDATCNCYDNCQHYWECPCDF 60
 QY 61 KRVCIAELCKGRCPSPRGRECDCAQCKYDKCCPDYESFCAE----- 106
 DB 61 KRVCIAELCKGRCPSPRGRECDCAQCKYDKCCPDYESFCAE----- 106
 QY 107 ----- 106
 DB 121 PFGASQTIKSTTKSRPKPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTI 180
 QY 107 -----VKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTPTDTST 147
 DB 181 KIKSSKNSAANRELQKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTPTDTST 240
 QY 148 TQHNKVSIPKLTAKINPRSLPNSDTSKETSITVNKETTVEKETTINKQTSIDG 207
 DB 241 TQHNKVSIPKLTAKINPRSLPNSDTSKETSITVNKETTVEKETTINKQTSIDG 300
 QY 208 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTKGALTPKEPTTTKPEPAS 267
 DB 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTKGALTPKEPTTTKPEPAS 360
 QY 268 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 327
 DB 361 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 420
 QY 328 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPTTTTPKPEAPTTKPEAPTT 387
 DB 421 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPTTTTPKPEAPTTKPEAPTT 480
 QY 388 EPAPTAPKAPPTTKPEAPTTKPEAPTTKPEPTTKPEAPTTKPEAPTTKSAPTTKPEAPT 447
 DB 481 EPAPTAPKAPPTTKPEAPTTKPEAPTTKPEPTTKPEAPTTKPEAPTTKSAPTTKPEAPT 540
 QY 448 TTKSAPTTKPEPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 507
 DB 541 TTKSAPTTKPEPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 600
 QY 508 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEEPTTT 567
 DB 601 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEEPTTT 660
 QY 568 PEEAPTTPKAAANPTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 627
 DB 661 PEEAPTTPKAAANPTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 720
 QY 628 APTTPKPEAPKELAPTTTKETSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKPE 687
 DB 721 APTTPKPEAPKELAPTTTKETSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKPE 780
 QY 688 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKPAETTPKETAETTPKPEAPTTPK 747

DB 781 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKPAETTPKETAETTPKPEAPTTPK 840
 QY 748 KPAPTTPETPTTSEVSTPTTTTKEPTTIHKSPDESTEPPELSABPTPKALENSPKPEGPT 807
 DB 841 KPAPTTPETPTTSEVSTPTTTTKEPTTIHKSPDESTEPPELSABPTPKALENSPKPEGPT 900
 QY 808 TKTPAATKPEMTTAKDKTERDLRTPPTTAAAPKMTKETATTTKTTESKITATTTQV 867
 DB 901 TKTPAATKPEMTTAKDKTERDLRTPPTTAAAPKMTKETATTTKTTESKITATTTQV 960
 QY 868 TSTTTQDTTTFKLTTLKTTTLPKVTTKKTIITTEIMNKPEETAKPKDRAATNSKATTPK 927
 DB 961 TSTTTQDTTTFKLTTLKTTTLPKVTTKKTIITTEIMNKPEETAKPKDRAATNSKATTPK 1020
 QY 928 PQKPTKAPKKPTSTKKPKTMPVRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 987
 DB 1021 PQKPTKAPKKPTSTKKPKTMPVRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1080
 QY 988 QTPNSKLVNPKSESDAGAGSETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1047
 DB 1081 QTPNSKLVNPKSESDAGAGSETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
 QY 1048 DETNICNGKPVDLGTLTLRNGTLVAFPHGYFWMLSPSPSPARRITEVWGIPSPIDTFT 1107
 DB 1141 DETNICNGKPVDLGTLTLRNGTLVAFPHGYFWMLSPSPSPARRITEVWGIPSPIDTFT 1200
 QY 1108 RCNCEGKTFKFDKSOYWFRTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPEV 1167
 DB 1201 RCNCEGKTFKFDKSOYWFRTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPEV 1260
 QY 1168 FPKRGSSIQOYIYKQEPVQKCGRRPALNYPYVGEMTQVRRRFRFAIGPSQHTTIRIQ 1227
 DB 1261 FPKRGSSIQOYIYKQEPVQKCGRRPALNYPYVGEMTQVRRRFRFAIGPSQHTTIRIQ 1320
 QY 1228 SPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYVND 1287
 DB 1321 SPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYVND 1380
 QY 1288 PSRTARAITRSGQTLKSVWYNCP 1311
 DB 1381 PSRTARAITRSGQTLKSVWYNCP 1404

RESULT 5

US-07-757-022B-62
 ; Sequence 62, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-62

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 1.8e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVTAEALCKGRCFESFERGECDCDAQCKYDKCCPDYEFSCAF----- 106
DB 61 KRVTAEALCKGRCFESFERGECDCDAQCKYDKCCPDYEFSCAFVHNPTSPSSKKAP 120
QY 107 ----- 106
DB 121 PPSGASQTKSTTKRSPKPNKKTKVLEBEITBEHSVSENQESSSSSSSSSTIW 180
QY 107 -----VKONKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 147
DB 181 KIKSSKNSAANRELQKKVKONKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 148 TQHNKVSSTGKITTAKPINRPSLPNSDTSKETSITVKNKEITVEKETTNNKQSTDG 207
DB 241 TQHNKVSSTGKITTAKPINRPSLPNSDTSKETSITVKNKEITVEKETTNNKQSTDG 300
QY 208 KEKTTSAKETOSTEKTSADKLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTPPKEPAS 267
DB 301 KEKTTSAKETQSLKTSADKLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTPPKEPAS 360
QY 268 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPE 327
DB 361 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPE 420
QY 328 APPTTKSAPTTKPEPAPTTKKAAPTTPKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 387
DB 421 APPTTKSAPTTKPEPAPTTKKAAPTTPKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTK 480
QY 388 EPAPTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTTKPEPAPT 447
DB 481 EPAPTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTTKPEPAPT 540
QY 448 TTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKPEPAPTTTKKP 507
DB 541 TTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKPEPAPTTTKKP 600
QY 508 APAPKPEPAPTTKPEPAPTTKPLTPTPEKLAPTTPKPAPTTPPELAPTTPEEPPTT 567
DB 601 APAPKPEPAPTTKPEPAPTTKPLTPTPEKLAPTTPKPAPTTPPELAPTTPEEPPTT 660
QY 568 PEEPAPTTPKAAAPNTKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKGTAPTTLKEP 627

DB 661 PEEPAPTTPKAAAPNTKPEPAPTTPKPEPAPTTPKPEPAPTTPKGTAPTTLKEP 720
QY 628 APPTPKKAPKELAPTTTKETSTTSDDKPAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPKG 687
DB 721 APPTPKKAPKELAPTTTKETSTTSDDKPAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPKG 780
QY 688 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDDKPAPTTPKETAPTTPKPEPAPTTPK 747
DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDDKPAPTTPKETAPTTPKPEPAPTTPK 840
QY 748 KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDSESTBELSAEPTPKALENSPKPEGVT 807
DB 841 KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDSESTBELSAEPTPKALENSPKPEGVT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLATTTPETTTAAAPKMKETATTTTEKTTESKITATTTQV 867
DB 901 TKTPAATKPEMTTAKDKTTERDLATTTPETTTAAAPKMKETATTTTEKTTESKITATTTQV 960
QY 868 TSTTTQDPTTTFKITTLLKTTLLAPKVTTTTKKTIITTEIMNKPETAKPKDRATNSKATTPK 927
DB 961 TSTTTQDPTTTFKITTLLKTTLLAPKVTTTTKKTIITTEIMNKPETAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKPTSTKKPTMPRVKPKTTPTPRKMSTMPPELNPTSRIAEAMLOTTTRN 987
DB 1021 POKPTKAPKPTSTKKPTMPRVKPKTTPTPRKMSTMPPELNPTSRIAEAMLOTTTRN 1080
QY 988 QTPNSKLEVEVNPKSDAGAGETPHMLLRPHVMEVETPDMDYLRVVPNOGIIINPMLS 1047
DB 1081 QTPNSKLEVEVNPKSDAGAGETPHMLLRPHVMEVETPDMDYLRVVPNOGIIINPMLS 1140
QY 1048 DETN1CNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1107
DB 1141 DETN1CNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1200
QY 1108 RNCSEKTPFFKDSQWRFRTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1167
DB 1201 RNCSEKTPFFKDSQWRFRTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1168 EFKRGGSIQOYIKQSPVOKPCRRPALNYPVVGEMTVRRRRFERAIGPSQTHTRIQY 1227
DB 1261 EFKRGGSIQOYIKQSPVOKPCRRPALNYPVVGEMTVRRRRFERAIGPSQTHTRIQY 1320
QY 1228 SPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
DB 1321 SPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
QY 1288 PSTARAITTRSGQTLISKVWYNCP 1311
DB 1381 PSTARAITTRSGQTLISKVWYNCP 1404

RESULT 6
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jav, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 1.8e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAE----- 106
Db 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAE VHNTPSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHVSSENQESSSSSSSSSTIR 180
QY 107 -----VKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTDG 300
QY 208 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTTPKEPAS 360
QY 268 TTPKGPPTTIKSAPTTPKEPATTTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTTKP 327
Db 361 TTPKGPPTTIKSAPTTPKEPATTTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTTKP 420
QY 328 APPTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 387
Db 421 APPTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 480
QY 388 EPAPTAPKKPAPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 447
Db 481 EPAPTAPKKPAPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 540
QY 448 TTKSAPTTPKGPSTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 507
Db 541 TTKSAPTTPKGPSTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 600
QY 508 APPTAPKEPATTTPKETAPTTPKLTTPTEKLAPTTPKEPATTTTPKEPATTTTPKEPATTTTPK 567
Db 601 APPTAPKEPATTTPKETAPTTPKLTTPTEKLAPTTPKEPATTTTPKEPATTTTPKEPATTTTPK 660
QY 568 PEEAPTTPKAAAPNTPKAPKELAPTTTKGPTSTTSKAPPTTPKEPATTTTPKEPATTTTPK 627
Db 661 PEEAPTTPKAAAPNTPKAPKELAPTTTKGPTSTTSKAPPTTPKEPATTTTPKEPATTTTPK 720
QY 628 APPTPKAPKELAPTTTKGPTSTTSKAPPTTPKGTAPTTPKEPATTTTPKEPATTTTPK 687
Db 721 APPTPKAPKELAPTTTKGPTSTTSKAPPTTPKGTAPTTPKEPATTTTPKEPATTTTPK 780
QY 688 TAPTTLKEPATTTPKAPKELAPTTTKGPTSTTSKAPPTTPKEPATTTTPKEPATTTTPK 747
Db 781 TAPTTLKEPATTTPKAPKELAPTTTKGPTSTTSKAPPTTPKEPATTTTPKEPATTTTPK 840
QY 748 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSPPDESTPELSAETPKALENSPKERGPT 807
Db 841 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSPPDESTPELSAETPKALENSPKERGPT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTQV 867
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTQV 960
QY 868 TSTTTQDTTPKINTLKTTLAPKVTTTKKTIITTEIMNKPEETAKPKORATNSKATTPK 927
Db 961 TSTTTQDTTPKINTLKTTLAPKVTTTKKTIITTEIMNKPEETAKPKORATNSKATTPK 1020
QY 928 PQKPTAPKKPTSTKPKTWPVRVKPTTTPPKWMTSTMPELNPTSRIAEAMLTQTTTRPN 987
Db 1021 PQKPTAPKKPTSTKPKTWPVRVKPTTTPPKWMTSTMPELNPTSRIAEAMLTQTTTRPN 1080

QY 988 QTPNSKLVENVNPKSEDAAGAGETPHMLLRPHVPMPEVTPDMDYLRPVNQGIINPMLS 1047
Db 1081 QTPNSKLVENVNPKSEDAAGAGETPHMLLRPHVPMPEVTPDMDYLRPVNQGIINPMLS 1140
QY 1048 DETNINCNGKPVDCGLTTLRNGTTLVAFRGHYFWMLSPPSPSPARRITEVWIGIPSPIDVFT 1107
Db 1141 DETNINCNGKPVDCGLTTLRNGTTLVAFRGHYFWMLSPPSPSPARRITEVWIGIPSPIDVFT 1200
QY 1108 RCNCEGKTFKKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTFKKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1168 FFKRGGSIOQYIYKQBPVQKCPGRRPALNYPVYGMTQVRRRFFERAIGESQTHTIRIQY 1227
Db 1261 FFKRGGSIOQYIYKQBPVQKCPGRRPALNYPVYGMTQVRRRFFERAIGESQTHTIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPGDYDYAFSKDQYINIDV 1287
Db 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPGDYDYAFSKDQYINIDV 1380
QY 1288 PSRTARAITRRSGQTLISKVWYNCP 1311
Db 1381 PSRTARAITRRSGQTLISKVWYNCP 1404

RESULT 7

US-10-164-595-78
; Sequence 78, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-78

Query Match 99.5%; Score 7029.7; DB 4; Length 1404;
Best Local Similarity 93.2%; Pred. No. 4.7e-192;
Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAE----- 106
Db 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAE VHNTPSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHVSSENQESSSSSSSSSTIR 180
QY 107 -----VKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTDG 300
QY 208 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTTPKEPAS 360
QY 268 TTPKGPPTTIKSAPTTPKEPATTTTKSAPTTPKEPATTTTPKEPATTTTPKEPATTTTKP 327

Db 224 EKTSADLAPTCKVLAKPTKAEITTKGPAITTKKEPTTTPKBPASTTPKEPTPTTIKS 283
QY 281 APTTPKEPAPTTTKSAPTTPKEPAPTTTKBPAPTTTKBPAPTTTKSAPTTPK 340
Db 284 APTTPKEPAPTTTKSAPTTPKEPAPTTTKBPAPTTTKBPAPTTTKSAPTTPK 343
QY 341 EPAPTTPKKAPPTTPKEPAPTTTPKEPTTTPKBPAPTTTKBPAPTTTKBPAPTT 400
Db 344 EPAPTTPKKAPPTTPKEPAPTTTPKEPTTTPKBPAPTTTKBPAPTTTKBPAPTT 403
QY 401 TPKEPAPTTKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPTKSAPTTPKEPS 460
Db 404 TPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPTKSAPTTPKEPS 463
QY 461 PTTTKBPAPTTTPKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTKBPAPTTTP 520
Db 464 PTTTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 523
QY 521 KETAPTTPKLPTTTPKEAPTTTPKEPAPTTPEELAPTTPEEPTTPPEPAPTTKAAA 580
Db 524 KETAPTTPKLPTTTPKEAPTTTPKEPAPTTPEELAPTTPEEPTTPPEPAPTTKAAA 583
QY 581 PNTPKBPAPTTTPKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 640
Db 584 PNTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 643
QY 641 APTTTKEPTSTSDKAPTTTPKGTATTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 700
Db 644 APTTTKEPTSTSDKAPTTTPKGTATTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 703
QY 701 PKKBPAPLAPTTTKGPTSTSDKAPTTTPKGTATTPKBPAPTTTPKBPAPTTTPKBPAPTT 760
Db 704 PKKBPAPLAPTTTKGPTSTSDKAPTTTPKGTATTPKBPAPTTTPKBPAPTTTPKBPAPTT 763
QY 761 TSEVSTPTTKPTTHKSPDSESTPESLAPTPKALNSPKGPGVPTTKPAATKPEMTT 820
Db 764 TSEVSTPTTKPTTHKSPDSESTPESLAPTPKALNSPKGPGVPTTKPAATKPEMTT 823
QY 821 TAKDKTTERDLPTTPTTAAKPKTKETATTTTEKTTESKITATTQVSTTTQDTPPKI 880
Db 824 TAKDKTTERDLPTTPTTAAKPKTKETATTTTEKTTESKITATTQVSTTTQDTPPKI 883
QY 881 TTLKTTTLAPKVTTKKIITTEIMNKPRETAKPKDRATNSKATTPKPKPTKPKKPTS 940
Db 884 TTLKTTTLAPKVTTKKIITTEIMNKPRETAKPKDRATNSKATTPKPKPTKPKKPTS 943
QY 941 TKPKTMRVRKPTTPTRKMTSTMPBLNPTSRIBABLQTTTPRNPOTPNKSLVEVNP 1000
Db 944 TKPKTMRVRKPTTPTRKMTSTMPBLNPTSRIBABLQTTTPRNPOTPNKSLVEVNP 1003
QY 1001 SEDAGGAEGETPHMLLRPHVFWPEVTPDMVYLPRVFNQIILINPMLSDETNCKGKPV 1060
Db 1004 SEDAGGAEGETPHMLLRPHVFWPEVTPDMVYLPRVFNQIILINPMLSDETNCKGKPV 1063
QY 1061 LTTTLNGLTLVAFRGHYFWMLSPFPSPSPARRITEVWGPISPIDIVTRCNCEGKTFPF 1120
Db 1064 LTTTLNGLTLVAFRGHYFWMLSPFPSPSPARRITEVWGPISPIDIVTRCNCEGKTFPF 1123
QY 1121 SQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAAALSTAKYKNWPSVFPFKGGISIQY 1180
Db 1124 SQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAAALSTAKYKNWPSVFPFKGGISIQY 1183
QY 1181 KOEPVQKCPGRRPALNPVYVEMTQVRRRRRERAIQPSQTHIRIQSPASLAYQDKGVL 1240
Db 1184 KOEPVQKCPGRRPALNPVYVEMTQVRRRRRERAIQPSQTHIRIQSPASLAYQDKGVL 1243
QY 1241 HNEVKVSLMRGLPNVVTSAISLNRKPDGYDYAFSKOQYINIDVPSRTARAITTRSG 1300
Db 1244 HNEVKVSLMRGLPNVVTSAISLNRKPDGYDYAFSKOQYINIDVPSRTARAITTRSG 1303
QY 1301 QTLKSKWVNCP 1311
|||||

Db 1304 QTLKSKWVNCP 1314
RESULT 9
US-07-757-022B-44
; Sequence 44, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Gesner, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-44
Query Match 96.3%; Score 6799.9; DB 4; Length 1270;
Best Local Similarity 96.9%; Pred. No. 1.4e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSLSCAGRCGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSLSCAGRCGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVCIAELCKGRCFESFEGRECDCCDAOCKKYDKCCPDYSPCAEVKDNKNRKKKPT 120
Db 26 -----ELSCKGRCFESFEGRECDCCDAOCKKYDKCCPDYSPCAEVKDNKNRKKKPT 79
QY 121 PKPVPVDEAGSLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKETINPRPSLPNPSDTSKE 180
Db 80 PKPVPVDEAGSLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKETINPRPSLPNPSDTSKE 139
|||||

QY	181	TSLVNKEITVETKETTITNTKQTSFSGKKEKTSKAKETOSIEKTSKAKOLAPTSKVLAKPTP	240
Db	140	TSLVNKEITVETKETTITNTKQTSFSGKKEKTSKAKETOSIEKTSKAKOLAPTSKVLAKPTP	199
QY	241	KAETTTKGPALITPKPEPTTTPKEBPASTPKPEPTTTPKSAPTTTPKEBPATTTKSAPTTTP	300
Db	200	KAETTTKGPALITPKPEPTTTPKEBPASTPKPEPTTTPKSAPTTTPKEBPATTTKSAPTTTP	259
QY	301	KEPAPTTTTPKEBPATTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTPKEPAP	360
Db	260	KEPAPTTTTPKEBPATTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTPKEPAP	319
QY	361	TTTPEPTTTPKEBPATTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTPKEPAP	420
Db	320	TTTPEPTTTPKEBPATTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTPKEPAP	379
QY	421	PSPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	480
Db	380	PSPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	439
QY	481	KKPAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	540
Db	440	KKPAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	499
QY	541	PTTPEKPAPTTPEELAPTTPEEPPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	600
Db	500	PTTPEKPAPTTPEELAPTTPEEPPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	559
QY	601	KEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	660
Db	560	KEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	619
QY	661	PKGTAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	720
Db	620	PKGTAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	679
QY	721	TSDKPAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	780
Db	680	TSDKPAPTTTPKEBPATTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKKEPAPTTTP	739
QY	781	DESTPELSAETPKALENSPKGVPTTTPKPAATTKPEMTTAKDKITERDLRTTPETTTA	840
Db	740	DESTPELSAETPKALENSPKGVPTTTPKPAATTKPEMTTAKDKITERDLRTTPETTTA	799
QY	841	APKMTKETATTTTESKITATTTQVTSITTTQVTSITTTQVTSITTTQVTSITTTQVTSIT	900
Db	800	APKMTKETATTTTESKITATTTQVTSITTTQVTSITTTQVTSITTTQVTSITTTQVTSIT	859
QY	901	TTTBMKPBETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKKPMRVRKPKTTTPTPR	960
Db	860	TTTBMKPBETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKKPMRVRKPKTTTPTPR	919
QY	961	KMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAAGAGETPHMLRPHV	1020
Db	920	KMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAAGAGETPHMLRPHV	979
QY	1021	FMEVPTPMDYLPVFNQGHIIINPMLSDNETNICNGKPDVGLTTLRNGTLVAFRGHYFWML	1080
Db	980	FMEVPTPMDYLPVFNQGHIIINPMLSDNETNICNGKPDVGLTTLRNGTLVAFRGHYFWML	1039
QY	1081	SPFSPSPARRITEVWGIIPSPIDVFTRCNCEGKTFEFGKOSQYWRFTNDIKDAGYKPIF	1140
Db	1040	SPFSPSPARRITEVWGIIPSPIDVFTRCNCEGKTFEFGKOSQYWRFTNDIKDAGYKPIF	1099
QY	1141	KGFGGLTGQIVAAALSTAKYKNWPSVFFFKRGGSIQQYIYKQEPVQKCPGRRPALNVPVY	1200
Db	1100	KGFGGLTGQIVAAALSTAKYKNWPSVFFFKRGGSIQQYIYKQEPVQKCPGRRPALNVPVY	1159
QY	1201	GEMTQVRRRRPERAIGSQTHITRIQYSPARLAYQDKGVLNHVKVSLIWRGLPNVVTSA	1260
Db	1160	GEMTQVRRRRPERAIGSQTHITRIQYSPARLAYQDKGVLNHVKVSLIWRGLPNVVTSA	1219
QY	1261	ISLPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTTRSGQTLKSVWYNCP	1311

Db 1220 ISLPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTTRSGQTLKSVWYNCP 1270

RESULT 10

US-07-757-022B-142

Sequence 142, Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 BridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:

LENGTH: 1313 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-142

Query Match 96.1%; Score 6785.6; DB 4; Length 1313;

Best Local Similarity 93.8%; Pred. No. 3.7e-185; Indels 84; Gaps 2;

Matches 1270; Conservative 0; Mismatches 0

QY 1 MAWKTLPIYLLILLSVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYDNCQHYMECCDF 60

Db 1 MAWKTLPIYLLILLSVFVIQQVSSQ----- 25

QY 61 KRVCUTABLCKGRCFESFERGREGCDCAQCKKYDKCCPDYEFCAE----- 106

Db 26 -----ELSCCKGRCFESFERGREGCDCAQCKKYDKCCPDYEFCAEHSVSENQESSSS 79

QY 107 -----VKDNKKNTTKKPTPKPPVVDAGSLDNGD 137

Db 80 SSSSSSTTIWIKSKNSANRELQKLVKDNKKNTTKKPTPKPPVVDAGSLDNGD 139

QY 138 FKVTPDTSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEITTVETKETT 197
Db 140 FKVTPDTSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEITTVETKETT 199
QY 198 TTNKQTSIDGKEKTSKETSQISIKTSKAKOIAPTSKVLAKPTPKAETTTKGPALTTTKEP 257
Db 200 TTNKQTSIDGKEKTSKETSQISIKTSKAKOIAPTSKVLAKPTPKAETTTKGPALTTTKEP 259
QY 258 TPPTPKPASTTPKEPTTTIKSAPTTKPEAPTTTTSAPTTTKEAPTTTKEAPTTTKEP 317
Db 260 TPPTPKPASTTPKEPTTTIKSAPTTTKEAPTTTTSAPTTTKEAPTTTKEAPTTTKEP 319
QY 318 EPAPTTTKEAPTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 377
Db 320 EPAPTTTKEAPTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 379
QY 378 TKSPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 437
Db 380 TKSPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 439
QY 438 PTTTKEAPTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 497
Db 440 PTTTKEAPTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 499
QY 498 EPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 557
Db 500 EPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 559
QY 558 TPEEPTPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 617
Db 560 TPEEPTPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 619
QY 618 GTAPTTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 677
Db 620 GTAPTTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 679
QY 678 KEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 737
Db 680 KEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 739
QY 738 PKEPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 797
Db 740 PKEPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 799
QY 798 NSPKPVGVTTKTAAKPEMTTAKDKTTERDLRTTPTTAAKPKMTKETAATTTTEKITE 857
Db 800 NSPKPVGVTTKTAAKPEMTTAKDKTTERDLRTTPTTAAKPKMTKETAATTTTEKITE 859
QY 858 SKITATTTQVSTTTQDTTTPFKITLKTTLAPKVTTTKKTTTTEIMNKPETAAPKPDOR 917
Db 860 SKITATTTQVSTTTQDTTTPFKITLKTTLAPKVTTTKKTTTTEIMNKPETAAPKPDOR 919
QY 918 ATNSKATTPKPKPTKAPKPTSTTKPKTMPVRVKPTTPPKMTSTMPELNPTSRIAE 977
Db 920 ATNSKATTPKPKPTKAPKPTSTTKPKTMPVRVKPTTPPKMTSTMPELNPTSRIAE 979
QY 978 AMLQTTTREPNTNPSKLVENPKSADAGAGETPHMLLRHVFVPEVTPDMVDYLPVYN 1037
Db 980 AMLQTTTREPNTNPSKLVENPKSADAGAGETPHMLLRHVFVPEVTPDMVDYLPVYN 1039
QY 1038 QGIIINPMLSDETNCGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITEVWG 1097
Db 1040 QGIIINPMLSDETNCGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITEVWG 1099
QY 1098 IPSPIDTFTTCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQI1VAALSTA 1157
Db 1100 IPSPIDTFTTCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQI1VAALSTA 1159
QY 1158 KYKNWPSVYFFKGGSIQOVIYKQEPVQKCPGRPALNYPVYGMQVRRRRFERAIGP 1217
Db 1160 KYKNWPSVYFFKGGSIQOVIYKQEPVQKCPGRPALNYPVYGMQVRRRRFERAIGP 1219

QY 1218 SQTHTIRIQSPARLAYQDKGVILNEVKVSLMRGLPNVVTSAISLPIRKPQDGYDYAF 1277
Db 1220 SQTHTIRIQSPARLAYQDKGVILNEVKVSLMRGLPNVVTSAISLPIRKPQDGYDYAF 1279
QY 1278 SKDOYINIDVPSRPARAITTTSQGTLSKQWYNCP 1311
Db 1280 SKDOYINIDVPSRPARAITTTSQGTLSKQWYNCP 1313
RESULT 11
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-46
Query Match 96.0%; Score 6784.9; DB 4; Length 1320;
Best Local Similarity 93.3%; Pred. No. 3.9e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLSVFIQOVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQOVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KRVCATLSCKGRCFESFERGRCDCDAQCKKYDKCCPDYESFCA----- 105

Db 26 -----ELSCKGRCFESPERGECDCDAQCKYDKCCDPDYSFCAEVHNPSPSSKKAP 79
Qy 106 -----EVKDNKNRRTKKKPTPKPPVVDGAG 130
Db 80 PPSGASQIKSTTKESPKNKKTKKVIKESIEITVKNKKNRRTKKKPTPKPPVVDGAG 139
Qy 131 SGLDNGDFKVTTPDSTTQHNKVSSTPKITAKPINRPSLPDSDTSKESISLTYNKETT 190
Db 140 SGLDNGDFKVTTPDSTTQHNKVSSTPKITAKPINRPSLPDSDTSKESISLTYNKETT 199
Qy 191 VETKETTNNKOTSDDGKEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 250
Db 200 VETKETTNNKOTSDDGKEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 259
Qy 251 LTTTKEPTTTPKEPASITTPKEPTPTTIKSAFTTPKEPAPTTTKGAPTTPKPAPTTTKE 310
Db 260 LTTTKEPTTTPKEPASITTPKEPTPTTIKSAFTTPKEPAPTTTKGAPTTPKPAPTTTKE 319
Qy 311 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 370
Db 320 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 379
Qy 371 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 430
Db 380 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 439
Qy 431 PTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 490
Db 440 PTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 499
Qy 491 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 550
Db 500 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 559
Qy 551 TPEELAPTTPEEPTTPPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 610
Db 560 TPEELAPTTPEEPTTPPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 619
Qy 611 TAPTTPKGATPPTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 670
Db 620 TAPTTPKGATPPTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 679
Qy 671 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 730
Db 680 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 739
Qy 731 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 790
Db 740 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 799
Qy 791 PTPKALENSPKPVGPTTKTAAKPEMTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAK 850
Db 800 PTPKALENSPKPVGPTTKTAAKPEMTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAKDPTTAK 859
Qy 851 TTEKTTESKITATTQVSTTTQDTPPKITLTKTTLAPKVTTKTITLTKTTLAPKVTTKTITLTKTTLAPKVTTKTITLTK 910
Db 860 TTEKTTESKITATTQVSTTTQDTPPKITLTKTTLAPKVTTKTITLTKTTLAPKVTTKTITLTKTTLAPKVTTKTITLTK 919
Qy 911 TAKPKDRATNSKATTPKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTST 970
Db 920 TAKPKDRATNSKATTPKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTST 979
Qy 971 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFMPEVTPDMD 1030
Db 980 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFMPEVTPDMD 1039
Qy 1031 YLPRVNPNGIILNMLSDETNLCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1090
Db 1040 YLPRVNPNGIILNMLSDETNLCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1099
Qy 1091 RITEVWGIPTSDITVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKDIFKGFGLGTQOI 1150
Db 1100 RITEVWGIPTSDITVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKDIFKGFGLGTQOI 1159

Qy 1151 VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1210
Db 1160 VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1219
Qy 1211 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRGLPNVVTSAISLPIRKP 1270
Db 1220 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRGLPNVVTSAISLPIRKP 1279
Qy 1271 GYDYAFSKDQYINIDVPSRTARATITTSQGLTSKVWYNCP 1311
Db 1280 GYDYAFSKDQYINIDVPSRTARATITTSQGLTSKVWYNCP 1320

RESULT 12

US-07-757-032B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

Query Match 96.0%; Score 6784.9; DB 4; Length 1320;
Best Local Similarity 93.3%; Pred. No. 3.9e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
Qy 1 MAWKTLPIYLILLISLVFVIQQVSSQDLSSCAGCGEYSRDTACNCDYNCQHMECCPDF 60

Db 1 NAWKTLPIYLLGLLSVFIQVSSQ----- 25
QY 61 KRVCITAEISCKGRCFESPERGECDCDAQCKYDKCCPDYSEFCA----- 105
Db 26 -----ELSCGRGCFESPERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 106 -----EVKDNKKNRITKKPTPKPPVDEAG 130
Db 80 PPGASOTIKSTTKRSPKPNKKTKKVBIESEEITEVKDNKKNRITKKPTPKPPVDEAG 139
QY 131 SGLDNGDFKVTPTDSTTOHNVKSTSPKIIITAKINPRPSLPNSDTSKETSILVKNKETT 190
Db 140 SGLDNGDFKVTPTDSTTOHNVKSTSPKIIITAKINPRPSLPNSDTSKETSILVKNKETT 199
QY 191 VETKEITTTNKQTSDDGKEKITSAKETQSIKTSKADLAPTSKVLAKETPKAEITTKGPA 250
Db 200 VETKEITTTNKQTSDDGKEKITSAKETQSIKTSKADLAPTSKVLAKETPKAEITTKGPA 259
QY 251 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 310
Db 260 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 319
QY 311 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 370
Db 320 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 379
QY 371 PKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 430
Db 380 PKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 439
QY 431 PTTTKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 490
Db 440 PTTTKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 499
QY 491 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 550
Db 500 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 559
QY 551 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 610
Db 560 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 619
QY 611 TAPTTPKGAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 670
Db 620 TAPTTPKGAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 679
QY 671 EPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 730
Db 680 EPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 739
QY 731 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 790
Db 740 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 799
QY 791 PTPKALNSPKBPGVPTTKTAAKPEMTTAAKKTTERDLRTTPTTTTAAKMTKETAT 850
Db 800 PTPKALNSPKBPGVPTTKTAAKPEMTTAAKKTTERDLRTTPTTTTAAKMTKETAT 859
QY 851 TTEKTTESKITAATTTQVSTTODTTPPKITTLKTTTLAPKVTTTKTTTTEIMNKPEE 910
Db 860 TTEKTTESKITAATTTQVSTTODTTPPKITTLKTTTLAPKVTTTKTTTTEIMNKPEE 919
QY 911 TAKPKDRATNSKATTPKQKPTKAPKKTSTSKPKTTPRVRPKTTPRKTWTSTMPBLN 970
Db 920 TAKPKDRATNSKATTPKQKPTKAPKKTSTSKPKTTPRVRPKTTPRKTWTSTMPBLN 979
QY 971 PTSRIABAMLQTTTPNOTPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPDMD 1030
Db 980 PTSRIABAMLQTTTPNOTPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1031 YLPRVNOGIIINPMLSDETNICNGKPVDTTLNGLTVAFRGHYFWMLSPFPSPSPAR 1090

Db 1040 YLPRVNOGIIINPMLSDETNICNGKPVDTTLNGLTVAFRGHYFWMLSPFPSPSPAR 1099
QY 1091 RITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGOI 1150
Db 1100 RITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGOI 1159
QY 1151 VAALSTAKYKNWPSVFFKRGSGIOQYIKQBPVQKCPGRRPALNYPVYGEMTQVRRRR 1210
Db 1160 VAALSTAKYKNWPSVFFKRGSGIOQYIKQBPVQKCPGRRPALNYPVYGEMTQVRRRR 1219
QY 1211 FERAIGSQTHTRIQYSPARLAYQDKGVHLNHEVKSILWRGLPNNVTSAISLNPRIKPD 1270
Db 1220 FERAIGSQTHTRIQYSPARLAYQDKGVHLNHEVKSILWRGLPNNVTSAISLNPRIKPD 1279
QY 1271 GYDYAFSKQYNNIDVPSRTARAITTRSGQTLSSKWYNCP 1311
Db 1280 GYDYAFSKQYNNIDVPSRTARAITTRSGQTLSSKWYNCP 1320

RESULT 13

US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52

Query Match 96.0%; Score 6780.6; DB 4; Length 1363;
Best Local Similarity 90.5%; Pred. No. 5.5e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25

QY 61 KRVTAEALSCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE----- 106
Db 26 -----ELSCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE----- 79

QY 107 ----- 106

Db 80 PPGASQTIKSTTKRSPKPNKKTKKVIESBEITEHSVSNQESSSSSSSSSTIW 139

QY 107 -----VKONKNRTKKKTPKPPVVDEAGSLDNGDFKVTTPDTST 147

Db 140 KITSSKNSAANRELQKLLKVKONKNRTKKKTPKPPVVDEAGSLDNGDFKVTTPDTST 199

QY 148 TOHNVSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQSTDG 207

Db 200 TOHNVSTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETTVEKTTTNKQSTDG 259

QY 208 KKTTSKETSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPKBPAS 267

Db 260 KKTTSKETSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPKBPAS 319

QY 268 TTPKEBPTTIKSAPTTPKBPATTTTKSAPTTPKBPATTTTKBPATTTKEPATTTTKEP 327

Db 320 TTPKEBPTTIKSAPTTPKBPATTTTKSAPTTPKBPATTTTKBPATTTKEPATTTTKEP 379

QY 328 APTTTKSAPTTPKBPATTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 387

Db 380 APTTTKSAPTTPKBPATTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 439

QY 388 EPAPTAKKBPATTPKBPATTPKBPATTTKBPSTTPKBPATTTKSAPTTTKBPATTPK 447

Db 440 EPAPTAKKBPATTPKBPATTPKBPATTTKBPSTTPKBPATTTKSAPTTTKBPATTPK 499

QY 448 TTKSAPTTPKBPSTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 507

Db 500 TTKSAPTTPKBPSTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 559

QY 508 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 567

Db 560 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 619

QY 568 PEPAPATTPKAAAPNTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 627

Db 620 PEPAPATTPKAAAPNTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 679

QY 628 APTTPKBPAPKELAPTTTKBPTSTTSKBPATTPKGTAPTTPKBPATTPKBPATTPKBPATTPK 687

Db 680 APTTPKBPAPKELAPTTTKBPTSTTSKBPATTPKGTAPTTPKBPATTPKBPATTPKBPATTPK 739

QY 688 TAPTTLKEBPATTPKBPAPKSLAPTTTKGPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 747

Db 740 TAPTTLKEBPATTPKBPAPKSLAPTTTKGPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPK 799

QY 748 KPAPTTPBPTTPPTTSEVSTPTTKBPTTIHKSPDESTPELSAETPKALENSPKBPAGVPT 807

Db 800 KPAPTTPBPTTPPTTSEVSTPTTKBPTTIHKSPDESTPELSAETPKALENSPKBPAGVPT 859

QY 808 TKTPAAKPEMTTAAKKTBRDLATTPPETTTAAAPKMTKETATTTKETSKITATTTQV 867

Db 860 TKTPAAKPEMTTAAKKTBRDLATTPPETTTAAAPKMTKETATTTKETSKITATTTQV 919

QY 868 TSTTTQDTPPKITLTKITTLAPKVTTKKITTITTEIMNKPEETAKPKDRATNSKATTPK 927

Db 920 TSTTTQDTPPKITLTKITTLAPKVTTKKITTITTEIMNKPEETAKPKDRATNSKATTPK 979

QY 928 PQKPTKAPKPKPTSTKKPKTMPVRPKPTTTPRKMSTMPELNPTSRIAEAMLQTTTRPN 987

Db 980 PQKPTKAPKPKPTSTKKPKTMPVRPKPTTTPRKMSTMPELNPTSRIAEAMLQTTTRPN 1039

QY 988 QTPNSKLVENVNPKSBDAGAGETPHMLLRHVFHVEVTPDMDYLRPVNQGIIINPMLS 1047

Db 1040 QTPNSKLVENVNPKSBDAGAGETPHMLLRHVFHVEVTPDMDYLRPVNQGIIINPMLS 1099

QY 1048 DETNICNKGKPDVGLTTLRNGTLVAPRGHVFWMLSPPSPARRITEVWGIPSPIDTVP 1107

Db 1100 DETNICNKGKPDVGLTTLRNGTLVAPRGHVFWMLSPPSPARRITEVWGIPSPIDTVP 1159

QY 1108 RCNCEGKTFPFKDSQWRFPTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1167

Db 1160 RCNCEGKTFPFKDSQWRFPTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1219

QY 1168 PFKRGGSIOQYIYKQEPVQKCPGRPALNYPVYVGEVTVRRRRFERRAIGPSOHTTIRIOY 1227

Db 1220 PFKRGGSIOQYIYKQEPVQKCPGRPALNYPVYVGEVTVRRRRFERRAIGPSOHTTIRIOY 1279

QY 1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKQYYNIDV 1287

Db 1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKQYYNIDV 1339

QY 1288 PSTARAITRSGQTLISKVWYNCP 1311

Db 1340 PSTARAITRSGQTLISKVWYNCP 1363

RESULT 14

US-10-164-595-58

; Sequence 58; Application US/10164595

; Patent No. 6657054

; GENERAL INFORMATION:

; APPLICANT: Origene Technologies, Inc

; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides

; FILE REFERENCE: IU 103 R1

; CURRENT APPLICATION NUMBER: US/10/164,595

; CURRENT FILING DATE: 2002-06-10

; NUMBER OF SEQ ID NOS: 80

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 58

; LENGTH: 1320

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-164-595-58

Query Match 95.8%; Score 6769.9; DB 4; Length 1320;
Best Local Similarity 93.1%; Pred. No. 1.1e-184;
Matches 1267; Conservative 0; Mismatches 3; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25

QY 61 KRVTAEALSCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE----- 105

Db 26 -----ELSCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE----- 79

QY 106 -----EVDKNKNRTKKKTPKPPVVDEAG 130

Db 80 PPGASQTIKSTTKRSPKPNKKTKKVIESBEITEVDKNKNRTKKKTPKPPVVDEAG 139

QY 131 SGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETT 190

Db 140 SGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINPRPSLPNSDTSKETSLSLVNKKETT 199

QY 191 VETKETTITTNKQSTDGKEKTTSAKETQSIKTSIKTSAKOLAPTSKVLAKPTPKAETTTKGA 250

Db 200 VETKETTITTNKQSTDGKEKTTSAKETQSIKTSIKTSAKOLAPTSKVLAKPTPKAETTTKGA 259

QY 251 LTTTPKEBPTTPKBPASITPKBPTTIKSAPTTPKBPATTTTKSAPTTPKBPATTTTKE 310

Db 260 LTTTPKEBPTTPKBPASITPKBPTTIKSAPTTPKBPATTTTKSAPTTPKBPATTTTKE 319


```

QY 311 PAPTTPKEPAPTTTKSAPTTPKEPAPTTPKKAPATTTPKEPAPTTTPKEPTTT 370
Db |||||
QY 320 PAPTTPKEPAPTTTKSAPTTPKEPAPTTPKKAPATTTPKEPAPTTTPKEPTTT 379
Db |||||
QY 371 PKEPAPTTKEPAPTTKEPAPTTAKKAPATTTPKEPAPTTTPKEPAPTTTKESPPTPKPEPA 430
Db |||||
QY 380 PKEPAPTTKEPAPTTTPKEPAPTTAPKAPATTTPKEPAPTTTPKEPAPTTTKESPPTPKPEPA 439
Db |||||
QY 431 PTTTKSAPTTPKEPAPTTTKSAPTTPKEPSTTKKEPAPTTTPKEPAPTTTPKAPATTTPKE 490
Db |||||
QY 440 PTTTKSAPTTPKEPAPTTTKSAPTTPKEPSTTKKEPAPTTTPKEPAPTTTPKAPATTTPKE 499
Db |||||
QY 491 PAPTTPKEPAPTTTKKAPATPKAPKEPAPTTPKETAPTTPKLTPTTPKEKLAAPTTPKEKAPT 550
Db |||||
QY 500 PAPTTPKEPAPTTTKKAPATPKAPKEPAPTTPKETAPTTPKLTPTTPKEKLAAPTTPKEKAPT 559
Db |||||
QY 551 TPBELAPTTEPTPTTPPEPAPTTTPKAPANTPKKAPANTPKKAPANTPKKAPANTPKKAPANTPK 610
Db |||||
QY 560 TPBELAPTTEPTPTTPPEPAPTTTPKAAAPNTPKKAPANTPKKAPANTPKKAPANTPKKAPANTPK 619
Db |||||
QY 611 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTTPK 670
Db |||||
QY 620 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTTPK 679
Db |||||
QY 671 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPK 730
Db |||||
QY 680 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTTPK 739
Db |||||
QY 731 KETAPTTPKEPAPTTPKKAPATTTPPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 790
Db |||||
QY 740 KETAPTTPKEPAPTTPKKAPATTTPPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 799
Db |||||
QY 791 PTPKALENPKKPGVPTTKTAPATKPEMTTAKDKTERDLRTPETTTAAPTWTETAT 850
Db |||||
QY 800 PTPKALENPKKPGVPTTKTAPATKPEMTTAKDKTERDLRTPETTTAAPTWTETAT 859
Db |||||
QY 851 TTEKTTESKITATTQVTSITTTQDTTPFKITTLTKTTLAPKVTTKTKTITTIHKMKPER 910
Db |||||
QY 860 TTEKTTESKITATTQVTSITTTQDTTPFKITTLTKTTLAPKVTTKTKTITTIHKMKPER 919
Db |||||
QY 911 TAKPKDRATNSKATTPKPKAPKAPKPKTSTKKPKTMVRKPKTTTPTRKMTSIMPEN 970
Db |||||
QY 920 TAKPKDRATNSKATTPKPKAPKAPKPKTSTKKPKTMVRKPKTTTPTRKMTSIMPEN 979
Db |||||
QY 971 PPSRAEAMLOTTTRPNQTPNSKLVVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMD 1030
Db |||||
QY 980 PPSRAEAMLOTTTRPNQTPNSKLVVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMD 1039
Db |||||
QY 1031 YLPRVNOGIIINPMLSDETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMVLSPPSPSPAR 1090
Db |||||
QY 1040 YLPRVNOGIIINPMLSDETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMVLSPPSPSPAR 1099
Db |||||
QY 1091 RITEVWGIPSPIDTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKIPFGFGLTGOI 1150
Db |||||
QY 1100 RITEVWGIPSPIDTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKIPFGFGLTGOI 1159
Db |||||
QY 1151 VAALSTAKYKNWPESVYFFKGGSIQQYIYKQEPVQKCPGRPALNYPVYGETTQVRRRR 1210
Db |||||
QY 1160 VAALSTAKYKNWPESVYFFKGGSIQQYIYKQEPVQKCPGRPALNYPVYGETTQVRRRR 1219
Db |||||
QY 1211 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWGLPNVVTSAISLPIRKPD 1270
Db |||||
QY 1220 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWGLPNVVTSAISLPIRKPD 1279
Db |||||
QY 1271 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1311
Db |||||
QY 1280 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1320
Db |||||

```

RESULT 15

US-07-757-022B-104

; Sequence 104, Application US/07757022B

```

; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

```

Query Match 79.3%; Score 5603.7; DB 4; Length 1140;
 Best Local Similarity 91.8%; Pred. No. 1.3e-151;
 Matches 1047; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

```

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDTACNDYNCQHYMECCPDF 60
Db |||||
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDTACNDYNCQHYMECCPDF 60
Db |||||
QY 61 KRVTCTAELSCKRCFCSEFERGECDCDAQCKKYDKCCPDYSPCAE----- 106
Db |||||
QY 61 KRVTCTAELSCKRCFCSEFERGECDCDAQCKKYDKCCPDYSPCAEVEHNPTSPSSKKAP 120
Db |||||
QY 107 ----- 106
Db |||||
QY 121 PPGASQTIKSTTKRSPKPNKKTKKVIIESEITEHVSSENQESSSSSSSSSSSIW 180
Db |||||
QY 107 -----VKDNKKRNTKKKTPKPPVDEAGSLDNGDPKVTTPDTST 147
Db |||||
QY 181 KIKSSKNSAANRELQKKLVKDNKKNTKKKTPKPPVDEAGSLDNGDKFVTTPDTST 240
Db |||||
QY 148 TOHNKVTSPKITAKDINPRPSLPNSDTSKETSLTWNKETTVEKTTTNNKQSTGDG 207
Db |||||

```

241	Db	TOHNKVSTGPKITIAKPIINPRPSLPNSDTSKETSUVNKEVTVETKEITNTTNKQTSIDG	300
208	Qy	KEKTSAKETQSTKTSKADLAPTSKVLAKPPTPKAETTTKGPAITTPKEPTPTTPKEPAS	267
301	Db	KEKTSAKETQSTKTSKADLAPTSKVLAKPPTPKAETTTKGPAITTPKEPTPTTPKEPAS	360
268	Qy	TTTKEPTPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	327
361	Db	TTTKEPTPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	420
328	Qy	APTITTKSAPTTTKEPAPTTTPKKAPATTTPKEPAPTTTPKEPTPTTPKEPAPTTTKEPAPTTPK	387
421	Db	APTITTKSAPTTTKEPAPTTTPKKAPATTTPKEPAPTTTPKEPTPTTPKEPAPTTTKEPAPTTPK	480
388	Qy	EPAPTAPKKAPATTPPKPAPATTTPKEPAPTTTKGSPSTTPKEPAPTTTKEPAPTTTKEPAPT	447
481	Db	EPAPTAPKKAPATTPPKPAPATTTPKEPAPTTTKGSPSTTPKEPAPTTTKEPAPTTTKEPAPT	540
448	Qy	TTKSAPTTTKEPSPSTTTTKEPAPTTTPKEPAPTTTPKKAPATTTPKEPAPTTTKEPAPTTPKKP	507
541	Db	TTKSAPTTTKEPSPSTTTTKEPAPTTTPKEPAPTTTPKKAPATTTPKEPAPTTTKEPAPTTPKKP	600
508	Qy	APTAPKEPAPTTPKETAPTTPKKLTPTTTPBKLAPTTTPKEPAPTTTPBELAPTTTPBEPTPT	567
601	Db	APTAPKEPAPTTPKETAPTTPKKLTPTTTPBKLAPTTTPKEPAPTTTPBELAPTTTPBEPTPT	660
568	Qy	PEEPAPTTTPKAAANPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTTLKEP	627
661	Db	PEEPAPTTTPKAAANPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTTLKEP	720
628	Qy	APTTPKKAPKELAPTTTKEPTSTTSOKPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKG	687
721	Db	APTTPKKAPKELAPTTTKEPTSTTSOKPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKG	780
688	Qy	TAPTTLKEPAPTTTPKKAPKELAPTTTKGPTSTTSDPAPTTTPKETAPTTPKETAPTTPKETAPTTPK	747
781	Db	TAPTTLKEPAPTTTPKKAPKELAPTTTKGPTSTTSDPAPTTTPKETAPTTPKETAPTTPKETAPTTPK	840
748	Qy	KCAPTTTTPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKPEGVPT	807
841	Db	KCAPTTTTPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKPEGVPT	900
808	Qy	TKTPAATKPEMTTTAKDKITTBEDURTTTPEITTAAPKMTKETAATTTTEKITSKITATTQV	867
901	Db	TKTPAATKPEMTTTAKDKITTBEDURTTTPEITTAAPKMTKETAATTTTEKITSKITATTQV	960
868	Qy	TSHTTQDPTTPFKITTLKITTLAPKVTTTKKITTTTEIMNKPEETAEPKPKDAATNSKATTPK	927
961	Db	TSHTTQDPTTPFKITTLKITTLAPKVTTTKKITTTTEIMNKPEETAEPKPKDAATNSKATTPK	1020
928	Qy	POKETKAPKKPTSTKKPKTMAPVRPKPTTTPPKWMTSTMPBLNPTSTRIAEAMLQTTTRPN	987
1021	Db	POKETKAPKKPTSTKKPKTMAPVRPKPTTTPPKWMTSTMPBLNPTSTRIAEAMLQTTTRPN	1080
988	Qy	QTPNSKLVENVPKSDEAGABGETPHMLLRPHVFPVPTTPWDVYLPRVPGNGIILNPMLS	1047
1081	Db	QTPNSKLVENVPKSDEAGABGETPHMLLRPHVFPVPTTPWDVYLPRVPGNGIILNPMLS	1140

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 133.379 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKVWVNC 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7064	100.0	1311	13	US-10-124-557-42
2	7049.7	99.8	1354	13	US-10-124-557-48
3	7049	99.8	1361	13	US-10-124-557-40
4	7044.7	99.7	1404	11	US-09-802-207-30
5	7044.7	99.7	1404	11	US-09-897-188-1
6	7044.7	99.7	1404	13	US-10-124-557-2
7	7044.7	99.7	1404	13	US-10-124-557-62
8	6811.3	96.4	1314	13	US-10-124-557-50
9	6799.9	96.3	1270	13	US-10-124-557-44
10	6785.6	96.1	1313	13	US-10-124-557-142
11	6784.9	96.0	1320	13	US-10-124-557-46
12	6784.9	96.0	1320	13	US-10-124-557-60
13	6780.6	96.0	1363	13	US-10-124-557-52
14	5603.7	79.3	1140	13	US-10-124-557-104

```

15 5500.7 77.9 1038 13 US-10-124-557-74
16 5472.6 77.5 1022 13 US-10-124-557-84
17 5344.6 75.7 1049 13 US-10-124-557-58
18 5011 70.9 941 13 US-10-124-557-14
19 3680.5 52.1 792 9 US-09-802-207-27
20 2850.9 40.4 538 14 US-10-038-694-3
21 2057.8 29.1 463 13 US-10-124-557-54
22 1806.8 25.6 423 13 US-10-124-557-66
23 1793.7 25.4 422 13 US-10-124-557-68
24 1723.1 24.4 372 13 US-10-124-557-64
25 1714.7 24.3 401 9 US-09-802-207-29
26 1398.2 19.8 5179 9 US-09-922-217-1068
27 1398.2 19.8 5179 9 US-09-833-263-1068
28 1398.2 19.8 5179 9 US-10-025-380-1068
29 1398.2 19.8 5179 16 US-10-734-564-121
30 1307.1 18.5 292 16 US-10-468-910-4
31 1219.6 17.3 237 13 US-10-124-557-72
32 1108.5 15.7 296 13 US-10-124-557-70
33 1025 14.5 188 14 US-10-038-694-2
34 969.9 13.7 1460 14 US-10-295-027-428
35 956.2 13.5 1367 9 US-09-801-368-108
36 951.7 13.5 3507 14 US-10-369-493-5784
37 945 13.4 6642 14 US-10-369-493-5013
38 941.5 13.3 19723 15 US-10-084-846A-5
39 931.3 13.2 1325 9 US-09-864-761-35612
40 884 12.5 157 13 US-10-124-557-102
41 884 12.5 157 13 US-10-124-557-114
42 879.4 12.4 5935 14 US-10-243-243A-8
43 877.1 12.4 5877 14 US-10-142-515-11
44 874.2 12.4 22152 16 US-10-715-066-5
45 873.5 12.4 2090 16 US-10-408-765A-2318

```

ALIGNMENTS

RESULT 1

```

US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

```

ATTORNEY/AGENT INFORMATION:
NAME: Ceerit, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: G1 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 100.0%; Score 7064; DB 13; Length 1311;
Best Local Similarity 100.0%; Pred. No. 3e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCYTAELSCGRCPESFERGECDCDAQCKYDKCCPDYSEFCAEVKONKRTKKPT 120
DB 61 KRVCYTAELSCGRCPESFERGECDCDAQCKYDKCCPDYSEFCAEVKONKRTKKPT 120

QY 121 KRPVVDVDEAGSLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180
DB 121 KRPVVDVDEAGSLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180

QY 181 TSLVTKETTVETKEITTTNKQISTDGEKTTSAKETQSIETKSADLAFTSKVLAKPTP 240
DB 181 TSLVTKETTVETKEITTTNKQISTDGEKTTSAKETQSIETKSADLAFTSKVLAKPTP 240

QY 241 KAETTTKGPALTTPKEBPTTPKEBASTTPKEBPTTPKSAPTTPKEBAPTTPKSAPTTP 300
DB 241 KAETTTKGPALTTPKEBPTTPKEBASTTPKEBPTTPKSAPTTPKEBAPTTPKSAPTTP 300

QY 301 KEAPATTTKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKKEBAPTTPKEBAPT 360
DB 301 KEAPATTTKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKKEBAPTTPKEBAPT 360

QY 361 TTPKEBPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKE 420
DB 361 TTPKEBPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKE 420

QY 421 PSPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 480
DB 421 PSPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 480

QY 481 KKPAPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 540
DB 481 KKPAPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 540

QY 541 PTTPEKAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 600
DB 541 PTTPEKAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 600

QY 601 KEAPATTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 660
DB 601 KEAPATTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 660

QY 661 PKGTAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 720
DB 661 PKGTAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 720

QY 721 TSDKAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 780
DB 721 TSDKAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 780

QY 781 DESTPELSAETPKALENSPKPGVPTTKTAAKPEMTTTAKDKTTERDLTTPETTTA 840

DB 781 DESTPELSAETPKALENSPKPGVPTTKTAAKPEMTTTAKDKTTERDLTTPETTTA 840

QY 841 APKMTKETATTEKTESKITATTTQVSTTTQDTPPKITTLTKTTTLAPKVTITTKKIT 900

DB 841 APKMTKETATTEKTESKITATTTQVSTTTQDTPPKITTLTKTTTLAPKVTITTKKIT 900

QY 901 TTEIMNKPEETAKPKDRATNSKATTPKQKPKTKAPKPKTSTTKPKTMPRVRKPKTTPTR 960

DB 901 TTEIMNKPEETAKPKDRATNSKATTPKQKPKTKAPKPKTSTTKPKTMPRVRKPKTTPTR 960

QY 961 KMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGAGSETPHMLLRPHV 1020

DB 961 KMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGAGSETPHMLLRPHV 1020

QY 1021 FMEVTPDMYLPVFNQGIINPMLSDETNINCCKPVDGLTTLNGLTLVAFRGHYFWML 1080

DB 1021 FMEVTPDMYLPVFNQGIINPMLSDETNINCCKPVDGLTTLNGLTLVAFRGHYFWML 1080

QY 1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQVWRFNTNDIKDAGYKPIF 1140

DB 1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQVWRFNTNDIKDAGYKPIF 1140

QY 1141 KGFGGLTGQIIVAALSTAKYNWPEVYFFKRGSSIIOYIYKQEPVQKCPGRRPALNYPV 1200

DB 1141 KGFGGLTGQIIVAALSTAKYNWPEVYFFKRGSSIIOYIYKQEPVQKCPGRRPALNYPV 1200

QY 1201 GEMTVQRRRRFERRAIGPSQTHIRIOYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSA 1260

DB 1201 GEMTVQRRRRFERRAIGPSQTHIRIOYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSA 1260

QY 1261 ISLPNIRKPDGYDYAFSDQYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1311

DB 1261 ISLPNIRKPDGYDYAFSDQYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1311

RESULT 2

US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserit, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match          99.8%; Score 7049.7; DB 13; Length 1354;
Best Local Similarity 96.8%; Pred. No. 6.9e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFEVIQQVSSQDLSSCAGRCGEGYSRDATCNCDCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFEVIQQVSSQDLSSCAGRCGEGYSRDATCNCDCYNCOHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCABEHSYSENQBSSSS 106
DB 61 KRVTAEISCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCABEHSYSENQBSSSS 120
QY 107 -----VKDNKKQRTKKKPTKPPVVDVDEAGSLDNGD 137
DB 121 SSSSSSTIWKIISKNKSAANRELQKKLVKDNKKQRTKKKPTKPPVVDVDEAGSLDNGD 180
QY 138 FKVTPTDSTQHNKYSTSPKITTAKINPRPSLPNSDTSKETSITVNKETTVEKETT 197
DB 181 FKVTPTDSTQHNKYSTSPKITTAKINPRPSLPNSDTSKETSITVNKETTVEKETT 240
QY 198 TTNKQISTDCKEKTSAKETQSIKTSKADLAPTSKVLAKEPTKAEITTKGPAITTPKEP 257
DB 241 TTNKQISTDCKEKTSAKETQSIKTSKADLAPTSKVLAKEPTKAEITTKGPAITTPKEP 300
QY 258 TPTPKPEASTTPKEPTTIKSAPTTKPEAPATTIKSAPTTKPEAPATTIKPEAPATTPK 317
DB 301 TPTPKPEASTTPKEPTTIKSAPTTKPEAPATTIKSAPTTKPEAPATTIKPEAPATTPK 360
QY 318 EPAPTTKEPAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 420
DB 361 EPAPTTKEPAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 437
QY 378 TKEPAPTTKEPAPTAPEKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 480
DB 421 TKEPAPTTKEPAPTAPEKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 497
QY 438 PTTTKEPAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 540
DB 481 PTTTKEPAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSA 557
QY 498 EPAPTTKKPAPTAPEKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 600
DB 541 EPAPTTKKPAPTAPEKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 617
QY 558 TPEEPTTTPEEAPTTKPAAPNTPKEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPK 660
DB 601 TPEEPTTTPEEAPTTKPAAPNTPKEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPK 677
QY 618 GTAPTTLKEPAPTTKPKAPKELAPTTTKEPTSTTSKAPATTGKGTAPTTKPEAPTT 720
DB 661 GTAPTTLKEPAPTTKPKAPKELAPTTTKEPTSTTSKAPATTGKGTAPTTKPEAPTT 737
QY 678 KEAPTTTKEPAPTTKPKAPKELAPTTTKEPTSTTSKAPATTGKGTAPTTKPEAPTT 780
DB 721 KEAPTTTKEPAPTTKPKAPKELAPTTTKEPTSTTSKAPATTGKGTAPTTKPEAPTT 797
QY 738 PKEAPTTPKKAPAPTTTPTPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALE 797

```

```

DB 781 PKBPAPTTPKKAPAPTTTPTPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALE 840
QY 798 NSPKPEGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKTE 857
DB 841 NSPKPEGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKTE 900
QY 858 SKITATTTQVSTTTQDTTTPFKITTLTKTTLAPKVTTKTKITTTIMNKPEETAKPKDR 917
DB 901 SKITATTTQVSTTTQDTTTPFKITTLTKTTLAPKVTTKTKITTTIMNKPEETAKPKDR 960
QY 918 ATNSKATTPKPKPTKAPKPTSTKKPKTPRVKPKTTPTPRMTSTMPELNPTSRAE 977
DB 961 ATNSKATTPKPKPTKAPKPTSTKKPKTPRVKPKTTPTPRMTSTMPELNPTSRAE 1020
QY 978 AMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPEHMLLRPHVFMPEVTPDMDYLPRVPN 1037
DB 1021 AMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPEHMLLRPHVFMPEVTPDMDYLPRVPN 1080
QY 1038 QGIIINPMLSDETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWG 1097
DB 1081 QGIIINPMLSDETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWG 1140
QY 1098 IPSPIDTFTVTRCNCCKTFFPKDSQYWRFTNDIKDAGYKPIPKGFGGLTGQIVAAALSTA 1157
DB 1141 IPSPIDTFTVTRCNCCKTFFPKDSQYWRFTNDIKDAGYKPIPKGFGGLTGQIVAAALSTA 1200
QY 1158 KYKNWPESVYFFKRGGSIOQYIYKQBPVQKCPGRRPALNVPVYGVMTQVRRRPFERAIGP 1217
DB 1201 KYKNWPESVYFFKRGGSIOQYIYKQBPVQKCPGRRPALNVPVYGVMTQVRRRPFERAIGP 1260
QY 1218 SQHTTIRIQVSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPIRKKPGDYDYAF 1277
DB 1261 SQHTTIRIQVSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPIRKKPGDYDYAF 1320
QY 1278 SKDOYNIIDVPSRTARAITTRSGQTLISKVWYNCP 1311
DB 1321 SKDOYNIIDVPSRTARAITTRSGQTLISKVWYNCP 1354

RESULT 3
US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990

```

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ceerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 99.8%; Score 7049; DB 13; Length 1361;
Best Local Similarity 96.3%; Pred. No. 7.2e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
1 MAWKTLPIVLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
1 MAWKTLPIVLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
61 KRVCIAELCKGRCFESFERGRCDDAQCKYDKCPCDYSEFCA----- 105
61 KRVCIAELCKGRCFESFERGRCDDAQCKYDKCPCDYSEFCAEVHNPTSPSSKKAP 120
106 -----EVKDKKNTKKKPKPPVVDDEAG 130
121 PPSGASQTIKSTKSPKPNKKTKVIESEBITEVKDKKNTKKKPKPPVVDDEAG 180
131 SGLDNGDFKVTTPDSTQHNVKYSTSPKIIITAKPINRPSLPPNSDTSKETSIVNKETT 190
181 SGLDNGDFKVTTPDSTQHNVKYSTSPKIIITAKPINRPSLPPNSDTSKETSIVNKETT 240
191 VETKEITTNKQSTDGKEKTTSAKETQSTIEKTSADLAPTSKVLAKPTPKAETTTKGPA 250
241 VETKEITTNKQSTDGKEKTTSAKETQSTIEKTSADLAPTSKVLAKPTPKAETTTKGPA 300
251 LTPPKETPTTPKEPASTTKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 310
301 LTPPKETPTTPKEPASTTKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 360
311 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPT 370
361 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPT 420
371 PKBPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEP 430
421 PKBPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEP 480
431 PTTTSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEP 490
481 PTTTSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEP 540
491 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 550
541 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 600
551 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEP 610
601 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEP 660
611 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 670
661 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 720
671 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 730

721 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 780
731 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 790
781 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 840
791 PTPKALENSPKPEPGVPTTKTAAATKPEMTTAAKDKTTERDLATTPETTTAAAPKMTKETAT 850
841 PTPKALENSPKPEPGVPTTKTAAATKPEMTTAAKDKTTERDLATTPETTTAAAPKMTKETAT 900
851 TTEKTTESKITATTTQVISTTTQDTPPKIITLKTTLAPKVTTKKTTTTEIMNKPE 910
901 TTEKTTESKITATTTQVISTTTQDTPPKIITLKTTLAPKVTTKKTTTTEIMNKPE 960
911 TAKPKDRATNSKATTPKPKPTKPKKPTSTKPKTTPRVRKPKTTPTRKMTSTMPELN 970
961 TAKPKDRATNSKATTPKPKPTKPKKPTSTKPKTTPRVRKPKTTPTRKMTSTMPELN 1020
971 PPSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFMVEVTPDMD 1030
1021 PPSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFMVEVTPDMD 1080
1031 YLPRVENQGIINPMLSDETNICNGKPVVDGLTTLRNGTLVAPRGHYFWMLSPEPSPAR 1090
1081 YLPRVENQGIINPMLSDETNICNGKPVVDGLTTLRNGTLVAPRGHYFWMLSPEPSPAR 1140
1091 RITEWGIISPDIPTVTRCNCEGKTFKDSQVWRFTNDIKDAGYKPKIPKFGGGLTGOI 1150
1141 RITEWGIISPDIPTVTRCNCEGKTFKDSQVWRFTNDIKDAGYKPKIPKFGGGLTGOI 1200
1151 VAALSTAKYKNWPEVYVFFKRGSGSIQYIYKQBPVQKCPGRPALNYPVVGEMTQVRRRR 1210
1201 VAALSTAKYKNWPEVYVFFKRGSGSIQYIYKQBPVQKCPGRPALNYPVVGEMTQVRRRR 1260
1211 FERAIGPSQTHITRIQYSPARLAYQKGVHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1270
1261 FERAIGPSQTHITRIQYSPARLAYQKGVHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1320
1271 GYDYAFSKDOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1311
1321 GYDYAFSKDOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1361

RESULT 4

US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US2002008624A1
; GENERAL INFORMATION:
; APPLICANT: Waxman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match 99.7%; Score 7044.7; DB 9; Length 1404;
Best Local Similarity 93.4%; Pred. No. 9.6e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVCCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVCCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVENHPTSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSKPPNKKTKKVIIESEEITEHSVSENQESSSSSSSSSIW 180
QY 107 -----VKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 147
Db 181 KIKSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNETKTTNKQSTSDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNETKTTNKQSTSDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360
QY 268 TTPKEPTPTTPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 327
Db 361 TTPKEPTPTTPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 420
QY 328 APPTTKSAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 387
Db 421 APPTTKSAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 480
QY 388 EPAPTAPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 447
Db 481 EPAPTAPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 540
QY 448 TTKSAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 507
Db 541 TTKSAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 600
QY 508 APAPTAPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 567
Db 601 APAPTAPKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 660
QY 568 PEPAPTTPKAAAPNTPKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 627
Db 661 PEPAPTTPKAAAPNTPKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 720
QY 628 APPTTKKAPKELAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 687
Db 721 APPTTKKAPKELAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 780
QY 688 TAPPTLKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 747
Db 781 TAPPTLKAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 840
QY 748 KPAPTTPPTTPPTSEVSTPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 807
Db 841 KPAPTTPPTTPPTSEVSTPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAPPTTKGAP 900
QY 808 TKTPAATKEMTTTAKDKTTERDLTTTPETTTAAAPKMTKETATTTKTESKITATTTQV 867
Db 901 TKTPAATKEMTTTAKDKTTERDLTTTPETTTAAAPKMTKETATTTKTESKITATTTQV 960
QY 868 TSTTTQDTPPKITTLKTTTLAPKVTTKKITTITTEIMNKBEETAKPKORATNSKATTPK 927
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKITTITTEIMNKBEETAKPKORATNSKATTPK 1020
QY 928 POKPTAPKPKPTSTKPKTPRVRKPKTTPPRKMTSTMPNLNPTSRFAEAMLOTTTPRN 987
Db 1021 POKPTAPKPKPTSTKPKTPRVRKPKTTPPRKMTSTMPNLNPTSRFAEAMLOTTTPRN 1080
QY 988 QTPNSKLVENVNPKBEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1047

Db 1081 QTPNSKLVENVNPKBEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
QY 1048 DETNICNGKPGVDGLTTLRNGTLVAFRGHYFWMILSPESPSPARRITEVWGPSPIDTVPST 1107
Db 1141 DETNICNGKPGVDGLTTLRNGTLVAFRGHYFWMILSPESPSPARRITEVWGPSPIDTVPST 1200
QY 1108 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1260
QY 1168 FFKEGGSIQQYIYKQEPVQKCPGRPALNIPVYGMTQVRRRRPERAIGPSQTHIRIQY 1227
Db 1261 FFKEGGSIQQYIYKQEPVQKCPGRPALNIPVYGMTQVRRRRPERAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGVHLNHNKVSILWRGLPNVVTSAISLNNIRKPDGYDYAFSKOQYINIDV 1287
Db 1321 SPARLAYQDKGVHLNHNKVSILWRGLPNVVTSAISLNNIRKPDGYDYAFSKOQYINIDV 1380
QY 1288 PSRTARAITTRSGQTLISKWYNCP 1311
Db 1381 PSRTARAITTRSGQTLISKWYNCP 1404
RESULT 5
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jav, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1
Query Match 99.7%; Score 7044.7; DB 11; Length 1404;
Best Local Similarity 93.4%; Pred. No. 9.6e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVCCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVCCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVENHPTSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSKPPNKKTKKVIIESEEITEHSVSENQESSSSSSSSSIW 180
QY 107 -----VKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 147
Db 181 KIKSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNETKTTNKQSTSDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNETKTTNKQSTSDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360

268 TTPKEPTTTIKSAPTTKEPAPTTTSGAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEP 327
Db
361 TTPKEPTTTIKSAPTTKEPAPTTTSGAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEP 420
Qy
328 APTTTKSAPTTKSPAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 387
Db
421 APTTTKSAPTTKSPAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 480
Qy
388 BEAPTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 447
Db
481 BEAPTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 540
Qy
448 TTKSAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 507
Db
541 TTKSAPTTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 600
Qy
508 APTAPKEPAPTTKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 567
Db
601 APTAPKEPAPTTKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 660
Qy
568 PEPAPTTKAAAPNTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 627
Db
661 PEPAPTTKAAAPNTKSPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720
Qy
628 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 687
Db
721 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780
Qy
688 TAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 747
Db
781 TAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 840
Qy
748 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 807
Db
841 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 900
Qy
808 TKPAPATKPEMTTAKDTERDLRTTPEPTTAAKPKMTKTAITTEKTESKITATTTOV 867
Db
901 TKPAPATKPEMTTAKDTERDLRTTPEPTTAAKPKMTKTAITTEKTESKITATTTOV 960
Qy
868 TSITTQDTPKTIITLKTTLAPKVTITTKITITTEIMNPEETAKPKDRATNSKATTPK 927
Db
961 TSITTQDTPKTIITLKTTLAPKVTITTKITITTEIMNPEETAKPKDRATNSKATTPK 1020
Qy
928 POKETKAPKPTSTKPKMTKPRVKPKTTTTPRKTMTSTMPBLNPTSRIAEAMLOTTTRPN 987
Db
1021 POKETKAPKPTSTKPKMTKPRVKPKTTTTPRKTMTSTMPBLNPTSRIAEAMLOTTTRPN 1080
Qy
988 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMVYLPVFNQGIINPMLS 1047
Db
1081 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMVYLPVFNQGIINPMLS 1140
Qy
1048 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITTEWGPSPIDIVFT 1107
Db
1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITTEWGPSPIDIVFT 1200
Qy
1108 RCNCEGKTFKKDSQYWFMTNDIKDAGYKPKIFKGFGLTQCIQAALSTAKYKWPESVY 1167
Db
1201 RCNCEGKTFKKDSQYWFMTNDIKDAGYKPKIFKGFGLTQCIQAALSTAKYKWPESVY 1260
Qy
1168 FFKRGGSITQYIKQEPVQKCPGRPALNYPVYGMTQVRRRRPERAIGPSQTHIRIQY 1227
Db
1261 FFKRGGSITQYIKQEPVQKCPGRPALNYPVYGMTQVRRRRPERAIGPSQTHIRIQY 1320
Qy
1228 SPARLAYODKGVLRHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYINIDV 1287
Db
1321 SPARLAYODKGVLRHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYINIDV 1380
Qy
1288 PSRTARAITTRSGQTLKSVYNCP 1311
Db
1381 PSRTARAITTRSGQTLKSVYNCP 1404

RESULT 6
US-10-124-557-2
; Sequence 2, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-124-557-2
Query Match 99.7%; Score 7044.7; DB 13; Length 1404;
Best Local Similarity 93.4%; Pred. No. 9.6e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 1 MAWKTLPIYLILLISLVFIQVSSQDLSSCAGRCGEGYSRDTACNDCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLILLISLVFIQVSSQDLSSCAGRCGEGYSRDTACNDCYNCOHYMECCPDF 60
Qy 61 KRVTAEISCKGRCFESFERGECDCDAQCKKYDKCCPDYSCFAE----- 106
Db 61 KRVTAEISCKGRCFESFERGECDCDAQCKKYDKCCPDYSCFAE----- 120
Qy 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEEITEHSVSENQESSSSSSSSSSSTIW 180
Qy 107 -----VRDNKNRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKLUKVDKNKNTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 148 TORNVSTSPKITTAKPINRPSLPNDSKETSITVKNKETTVEKTTTNNKQSTG 207
 Db 241 TORNVSTSPKITTAKPINRPSLPNDSKETSITVKNKETTVEKTTTNNKQSTG 300
 QY 208 KEXTSAKETQSEKTSKOLAPTSKVLAKPTKPAETTTKGPALTTKPEPTTTPKEPAS 267
 Db 301 KEXTSAKETQSEKTSKOLAPTSKVLAKPTKPAETTTKGPALTTKPEPTTTPKEPAS 360
 QY 268 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
 Db 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
 QY 328 APPTTKSAPTTKPEAPTTPKKAPTTKPEAPTTTKPEPTTTPKPEAPTTKPEAPTTPK 387
 Db 421 APPTTKSAPTTKPEAPTTPKKAPTTKPEAPTTTKPEPTTTPKPEAPTTKPEAPTTPK 480
 QY 388 EPAPTAPKPAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 447
 Db 481 EPAPTAPKPAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 540
 QY 448 TTKSAPTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 507
 Db 541 TTKSAPTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 600
 QY 508 APAPTAPKPAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 567
 Db 601 APAPTAPKPAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 660
 QY 568 PREAPPTPKAAAPNTKPEAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 627
 Db 661 PREAPPTPKAAAPNTKPEAPTTKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 720
 QY 628 APPTTKKAPKELAPTTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKPEAPT 687
 Db 721 APPTTKKAPKELAPTTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKPEAPT 780
 QY 688 TAPPTLKPEAPTTPKKAPKELAPTTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPT 747
 Db 781 TAPPTLKPEAPTTPKKAPKELAPTTTKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPT 840
 QY 748 KAPPTTTPPTTTPKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKPEAPT 807
 Db 841 KAPPTTTPPTTTPKPEPTTTPKPEAPTTPKKAPTTTKPEPTTTPKPEAPTTTKPEAPT 900
 QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKMTKETAATTEKTESKITATTTQV 867
 Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAKMTKETAATTEKTESKITATTTQV 960
 QY 868 TSTTTQDTPPKITTLKTTTLAPKVTTTKKIIITTEIMNKPDEATAKPKDRATNSKATTPK 927
 Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTTKKIIITTEIMNKPDEATAKPKDRATNSKATTPK 1020
 QY 928 POKPTKABKPTSTKPKTMRVKPKTTPPKMTSTWDELNPTSRJAEAMLQTTTRPN 987
 Db 1021 POKPTKABKPTSTKPKTMRVKPKTTPPKMTSTWDELNPTSRJAEAMLQTTTRPN 1080
 QY 988 QTPNSKLVNPKSEDAGAGETPHMLRLRHVFMPEVTPDMVYLPRVNOGIIINPMLS 1047
 Db 1081 QTPNSKLVNPKSEDAGAGETPHMLRLRHVFMPEVTPDMVYLPRVNOGIIINPMLS 1140
 QY 1048 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITEVWGPISPDITVFT 1107
 Db 1141 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITEVWGPISPDITVFT 1200
 QY 1108 RCNCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAAALSTAKYKNWPSVY 1167
 Db 1201 RCNCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAAALSTAKYKNWPSVY 1260
 QY 1168 FFKGGSIIQYIYKQEPVOKCPGRRPALNYPVYGEMTQVRRRFRERAIQPSQTHIRIQY 1227
 Db 1261 FFKGGSIIQYIYKQEPVOKCPGRRPALNYPVYGEMTQVRRRFRERAIQPSQTHIRIQY 1320

QY 1228 SPARLAYQDQGVILHNEVKVSIILMRGLPNVVTSAISLPIRKPBGYDYAFSKDQYVNDV 1287
 Db 1321 SPARLAYQDQGVILHNEVKVSIILMRGLPNVVTSAISLPIRKPBGYDYAFSKDQYVNDV 1380
 QY 1288 PSRTARAITTRSGQTILSKWYNCP 1311
 Db 1381 PSRTARAITTRSGQTILSKWYNCP 1404

RESULT 7
 US-10-124-557-62
 ; Sequence 62, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine C.
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Csert, Luana
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 62:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1404 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
 ; US-10-124-557-62

Query Match 99.7%; Score 7044.7; DB 13; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 9.6e-163;
 Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLILLISLVFIQQVSSQDLSSCAGRCGEGYSRDAICNDYNCQHYMECCPDF 60
 Db 1 MAWKTLPIYLLILLISLVFIQQVSSQDLSSCAGRCGEGYSRDAICNDYNCQHYMECCPDF 60
 QY 61 KRVCATLSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAE----- 106
 Db 61 KRVCATLSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKAP 120

Qy	107	-----	106
Db	121	PPSGASQTIKSTTKESPKNKKTKKVIIEEITEHSVSENQESSSSSSSSSTI	180
Qy	107	-----VKONKVRTKKPTPKPPVVDGAGSLDNGDFKVTTPDST	147
Db	181	KIKSSKNSAANRELQKLVKONKKNRKKKPTPKPPVVDGAGSLDNGDFKVTTPDST	240
Qy	148	TOHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNNKETTVTKETITNNKOTSTDG	207
Db	241	TOHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNNKETTVTKETITNNKOTSTDG	300
Qy	208	KEKTTSAKETOSIEKTSANDLAPTCKVLAKPTPKAETTTKGPAITPKPEPTTTPKBPAS	267
Db	301	KEKTTSAKETOSIEKTSANDLAPTCKVLAKPTPKAETTTKGPAITPKPEPTTTPKBPAS	360
Qy	268	TTPKPTPTTIKSAPTTPKEPAPTTPKSAPTTPKSPAPTTPKPEAPTTPKAPTTPKEP	327
Db	361	TTPKPTPTTIKSAPTTPKAPAPTTPKSAPTTPKSPAPTTPKPEAPTTPKAPTTPKEP	420
Qy	328	APTPTTKSAPTTPKEPAPTTPKAPAPTTPKSAPTTPKSPAPTTPKPEAPTTPKAPTTPK	387
Db	421	APTPTTKSAPTTPKEPAPTTPKAPAPTTPKSAPTTPKSPAPTTPKPEAPTTPKAPTTPK	480
Qy	388	EPAPTAPKKPAPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKAPTTPKEAPT	447
Db	481	EPAPTAPKKPAPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKAPTTPKEAPT	540
Qy	448	TTKSAPTTPKESPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKAPTTPKAP	507
Db	541	TTKSAPTTPKESPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKAPTTPKAP	600
Qy	508	APTAPKEPAPTTPKETAPTTPKKLTPTPEKLAPTTPKEPAPTTPPELAPTTPPEEPTPT	567
Db	601	APTAPKEPAPTTPKETAPTTPKKLTPTPEKLAPTTPKEPAPTTPPELAPTTPPEEPTPT	660
Qy	568	PEEPAPTTPKAAAPNTPKGAPTTPKEPAPTTPKPEAPTTPKAPTTPKAPTTPKAPTTPK	627
Db	661	PEEPAPTTPKAAAPNTPKGAPTTPKEPAPTTPKPEAPTTPKAPTTPKAPTTPKAPTTPK	720
Qy	628	APTTPKKAPKELAPTTPKETPTSTSDKAPTTPKGTAPTTPKEPAPTTPKPEAPTTPKG	687
Db	721	APTTPKKAPKELAPTTPKETPTSTSDKAPTTPKGTAPTTPKEPAPTTPKPEAPTTPKG	780
Qy	688	TAPTTLKEPAPTTPKKPAPKELAPTTPKGTSTSDKAPTTPKGTAPTTPKEAPTTPKAPTTPK	747
Db	781	TAPTTLKEPAPTTPKKPAPKELAPTTPKGTSTSDKAPTTPKGTAPTTPKEAPTTPKAPTTPK	840
Qy	748	KPAPTTPETPPPTSEVSTPTTTKEPTTIHKSPDESTPELSAPTTPKALENSPKPCVPT	807
Db	841	KPAPTTPETPPPTSEVSTPTTTKEPTTIHKSPDESTPELSAPTTPKALENSPKPCVPT	900
Qy	808	TKTPAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETAITTEKTTESKITATTQV	867
Db	901	TKTPAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETAITTEKTTESKITATTQV	960
Qy	868	TSITTTQDTPFKITTLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPK	927
Db	961	TSITTTQDTPFKITTLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPK	1020
Qy	928	POKPTKAPKPTSTKKKTPMRVKPTTTPRKTMTSTMPBLNPTSRIAEAMLOTTTRN	987
Db	1021	POKPTKAPKPTSTKKKTPMRVKPTTTPRKTMTSTMPBLNPTSRIAEAMLOTTTRN	1080
Qy	988	QTPNSKLVEVNPKESEDAGAGETPHMLLRHVFMPEVTPDMDYLPVPNQGIINPMLS	1047
Db	1081	QTPNSKLVEVNPKESEDAGAGETPHMLLRHVFMPEVTPDMDYLPVPNQGIINPMLS	1140
Qy	1048	DETNI CNKPKVDGLTTLNGLTVAFRGHYFWMLSFFSPSPSPARITEVWGIPSPIDTFT	1107
Db	1141	DETNI CNKPKVDGLTTLNGLTVAFRGHYFWMLSFFSPSPSPARITEVWGIPSPIDTFT	1200

1108 RCNCEGKTFPPKDSQYWRFTNDIKDAGYPKIFKFGGLTGQIVAAALSTAKYKNWPESVY 1167
1201 RCNCEGKTFPPKDSQYWRFTNDIKDAGYPKIFKFGGLTGQIVAAALSTAKYKNWPESVY 1260
1168 FFKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1227
1261 FFKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1320
1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYVAFSKDQYNNIDV 1287
1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYVAFSKDQYNNIDV 1380
1288 PSRTARAITTRSGQTLKSKWYNCP 1311
1381 PSRTARAITTRSGQTLKSKWYNCP 1404

RESULT 8
US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Geener, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match 96.4%; Score 6811.3; DB 13; Length 1314;
Best Local Similarity 96.0%; Pred. No. 4e-157;
Matches 1278; Conservative 4; Mismatches 12; Indels 37; Gaps 3;

QY 1 MAWKTLPIYLLILLUSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
Db |||||
QY 1 MAWKTLPIYLLILLUSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
Db |||||
QY 61 KRVCITAEULSC-----KRCFESFERGECDCDAQCKYDKCCPDY 100
Db |||||
QY 61 KRVCITAEHUSGENESSSSSSSSSTIWKIKSSKSAANRELQ-----KK----- 107
Db |||||
QY 101 ESFCAEVKDNKKNRTKKKPTKPPVWDEAGSLONGDFKVTPTDSTTOHNVSTSPKIT 160
Db |||||
QY 108 ----LKVKDNKKNRTKKKPTKPPVWDEAGSLONGDFKVTPTDSTTOHNVSTSPKIT 163
Db |||||
QY 161 TAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTSDGKEKTTSAKETQSI 220
Db |||||
QY 164 TAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTSDGKEKTTSAKETQSI 223
Db |||||
QY 221 EKTSAKDLAPTSKVLAKPTKAEITTKGPALTTPKEBPTTPPKBPASTTPKEBPTTPKS 280
Db |||||
QY 224 EKTSAKDLAPTSKVLAKPTKAEITTKGPALTTPKEBPTTPPKBPASTTPKEBPTTPKS 283
Db |||||
QY 281 APTTPKEBPTTPKSAPTTPKEBPTTPKBPATTPKEBPTTPKBPATTPKBPATTPK 340
Db |||||
QY 284 APTTPKEBPTTPKSAPTTPKEBPTTPKBPATTPKBPATTPKBPATTPKBPATTPK 343
Db |||||
QY 341 EPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 400
Db |||||
QY 344 EPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 403
Db |||||
QY 401 TPKEBPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 460
Db |||||
QY 404 TPKEBPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 463
Db |||||
QY 461 PTTTKEBPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 520
Db |||||
QY 464 PTTTKEBPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 523
Db |||||
QY 521 KETAPTTPKBLPTTPKEKLAPTTPKEBPTTPBELAPTTPPEBPTTPPEBAPTTPKAAA 580
Db |||||
QY 524 KETAPTTPKBLPTTPKEKLAPTTPKEBPTTPBELAPTTPPEBPTTPPEBAPTTPKAAA 583
Db |||||
QY 581 PNTKEBAPTTPKEBAPTTPKBPATTPKETAPTTPKGTAPTTPKBPATTPKBPAPKEL 640
Db |||||
QY 584 PNTKEBAPTTPKEBAPTTPKBPATTPKETAPTTPKGTAPTTPKBPATTPKBPAPKEL 643
Db |||||
QY 641 APTTTKEPTSTSDKPAPTTPKGTAPTTPKEBAPTTPKEBAPTTPKGTAPTTPKBPATTPK 700
Db |||||
QY 644 APTTTKEPTSTSDKPAPTTPKGTAPTTPKEBAPTTPKEBAPTTPKGTAPTTPKBPATTPK 703
Db |||||
QY 701 PKKPAPKELAPTTPKGTPTSDKPAPTTPKETAPTTPKEBAPTTPKBPATTPKBPATTPK 760
Db |||||
QY 704 PKKPAPKELAPTTPKGTPTSDKPAPTTPKETAPTTPKEBAPTTPKBPATTPKBPATTPK 763
Db |||||
QY 761 TSEVSTPTTPKEPTTHKSPDESTPELSAEPKPALENSPKBPBGVPTTKBPAATKPEMTT 820
Db |||||
QY 764 TSEVSTPTTPKEPTTHKSPDESTPELSAEPKPALENSPKBPBGVPTTKBPAATKPEMTT 823
Db |||||
QY 821 TAKDKTTERDLRTPTTPTTAAAPKMTKETAATTEKTTESKITAATTTQVSTTQDTPPKI 880
Db |||||
QY 824 TAKDKTTERDLRTPTTPTTAAAPKMTKETAATTEKTTESKITAATTTQVSTTQDTPPKI 883
Db |||||
QY 881 TTLKTTTLAPKVTTKKTIITTEIMNKPEETAAPKDRATNSKATTPKPKQKTPKPKKPTS 940
Db |||||
QY 884 TTLKTTTLAPKVTTKKTIITTEIMNKPEETAAPKDRATNSKATTPKPKQKTPKPKKPTS 943
Db |||||
QY 941 TKPKPTMVRVRKPTTPPKMTSTMPKNTSRIAEAMLQTTTRPNQTPNSKLVNPK 1000
Db |||||
QY 944 TKPKPTMVRVRKPTTPPKMTSTMPKNTSRIAEAMLQTTTRPNQTPNSKLVNPK 1003
Db |||||
QY 1001 SEDAGGAEGETPHMLLRPHVMEVTPDMDXLPRVNPQGIILNPMLSDETINCNGKVPDG 1060
Db |||||
QY 1004 SEDAGGAEGETPHMLLRPHVMEVTPDMDXLPRVNPQGIILNPMLSDETINCNGKVPDG 1063
Db |||||

QY 1061 LTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFTRCNCBCKTFFFKD 1120
Db |||||
QY 1064 LTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFTRCNCBCKTFFFKD 1123
Db |||||
QY 1121 SOYWRFTNDIKDAGYKPIFKGGLTGOIIVAALSTAKYKNWPESVYFFKGGSTQOYIY 1180
Db |||||
QY 1124 SOYWRFTNDIKDAGYKPIFKGGLTGOIIVAALSTAKYKNWPESVYFFKGGSTQOYIY 1183
Db |||||
QY 1181 KOEPVQKCPGRRPALNPVYGMTQVRRRRPERAIGPSQTHIRIOYSPARLAYODKGV 1240
Db |||||
QY 1184 KOEPVQKCPGRRPALNPVYGMTQVRRRRPERAIGPSQTHIRIOYSPARLAYODKGV 1243
Db |||||
QY 1241 HNEVKVILWRGLPNVVTSAISLNPINRKPQGYVYAFSKOQYNNIDVPSRTARAITTRSG 1300
Db |||||
QY 1344 HNEVKVILWRGLPNVVTSAISLNPINRKPQGYVYAFSKOQYNNIDVPSRTARAITTRSG 1303
Db |||||
QY 1301 QTLISKVWYNCP 1311
Db |||||
QY 1304 QTLISKVWYNCP 1314
Db |||||

RESULT 9

US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

1021 FMPEVTPMDYLPVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1080
980 FMPEVTPMDYLPVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1039
1081 SPFPSPSPARRITEVWGIPSPIDVFTTRCNCCEGKTFEFDKQYWFRTNDIKDAGYKPIF 1140
1040 SPFPSPSPARRITEVWGIPSPIDVFTTRCNCCEGKTFEFDKQYWFRTNDIKDAGYKPIF 1099
1141 KGFGGLTQIIVAAALSTAKYKNWPESVYFVKRGSGIQQYIYKQEPVQKCFGRPALNYPVY 1200
1100 KGFGGLTQIIVAAALSTAKYKNWPESVYFVKRGSGIQQYIYKQEPVQKCFGRPALNYPVY 1159
1201 GEMTOVRRRFRERAIQPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSA 1260
1160 GEMTOVRRRFRERAIQPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSA 1219
1261 ISLPIRKPDDGYDYAFSKDQYNNIDVPSRTARITTRSGOTLSKVWYNCP 1311
1220 ISLPIRKPDDGYDYAFSKDQYNNIDVPSRTARITTRSGOTLSKVWYNCP 1270

RESULT 10

US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth M.
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match 96.3%; Score 6799.9; DB 13; Length 1270;
Best Local Similarity 96.9%; Pred. No. 7.2e-157;
Matches 1270; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
1 MAWKTLPIYLLILLVFIQVSSQDLSACAGRCGEGYSDATCNCYNQCQHYMECCPDF 60
1 MAWKTLPIYLLILLVFIQVSSQ----- 25
61 KVCVTAELSCRCRCPESPERGECDCDAQCKYDKCCPDYEFCAEVKDNKNRKKKPT 120
26 -----ELSCRCRCPESPERGECDCDAQCKYDKCCPDYEFCAEVKDNKNRKKKPT 79
121 KPXPVVDEAGSLDGNDFKVTPTDSTTQHNVKSVSPKITTAKPINRPSLPPNSDTSKE 180
80 KPXPVVDEAGSLDGNDFKVTPTDSTTQHNVKSVSPKITTAKPINRPSLPPNSDTSKE 139
181 TSLTVNKETTVETKETTITNKQTSIDGKEKTTISAKETQSIKTSADLAPTSKVLAKPTP 240
140 TSLTVNKETTVETKETTITNKQTSIDGKEKTTISAKETQSIKTSADLAPTSKVLAKPTP 199
241 KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTP 300
200 KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTP 259
301 KEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 360
260 KEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 319
361 TTPKEPTPTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 420
320 TTPKEPTPTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 379
421 PSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 480
380 PSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 439
481 KKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 540
440 KKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 499
541 PTPPEKAPTTPEELAPTTPEEPPTTPPEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 600
500 PTPPEKAPTTPEELAPTTPEEPPTTPPEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP 559
601 KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 660
560 KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 619
661 PKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 720
620 PKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 679
721 TSDKAPATTTPKETAPTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 780
680 TSDKAPATTTPKETAPTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKETPTSTSDKAPATT 739
781 DESTPELSAETTPKALENSKPEPGVPTTKTAPKAPKELAPTTTKETPTSTSDKAPATT 840
740 DESTPELSAETTPKALENSKPEPGVPTTKTAPKAPKELAPTTTKETPTSTSDKAPATT 799
841 APKMTKETATTTKETSITATTTQVSTTTQDTPPKITLTKTTLLAPKVTITTKIT 900
800 APKMTKETATTTKETSITATTTQVSTTTQDTPPKITLTKTTLLAPKVTITTKIT 859
901 TTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPKTTPR 960
860 TTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPKTTPR 919
961 KWTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVENVNPKSDAGAGETPHMLLRPHV 1020
920 KWTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVENVNPKSDAGAGETPHMLLRPHV 979

Query Match			
Best Local Similarity 96.1%; Score 6785.6; DB 13; Length 1313;			
Matches 1270; Conservative 0; Mismatches 0; Indels 84; Gaps 2;			
QY	1	MAWKTLPYLLILLISLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF	60
Db	1	MAWKTLPYLLILLISLVFVIQVSSQ-----	25
QY	61	KRVCTABLSCKGRCFEFGREDCDCAQCKKYDKCCPDYESCAE-----	106
Db	26	-----ELSCKGRCFEFGREDCDCAQCKKYDKCCPDYESFCAEHSVSENQESSSS	79
QY	107	-----VKNKKNRITKKPTPKPPVVDVAGSGLDNGD	137
Db	80	SSSSSSSTIWIKISSKNSAANRELQKKLVKDKNKKNRITKKPTPKPPVVDVAGSGLDNGD	139
QY	138	FKVTPDTSITONHKVSTSKITIAKINPRPSLPNSDTSKETSITVNETKETT	197
Db	140	FKVTPDTSITONHKVSTSKITIAKINPRPSLPNSDTSKETSITVNETKETT	199
QY	198	TINKQSTDGKEKTSIAKETQSIEKTSKVLAKPTPKAETTTKGPALTTTPKEP	257
Db	200	TINKQSTDGKEKTSIAKETQSIEKTSKVLAKPTPKAETTTKGPALTTTPKEP	259
QY	258	TPPTPKPASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKPAPTTPK	317
Db	260	TPPTPKPASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKPAPTTPK	319
QY	318	EPAPTTPKEPAPTITTKSAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPTPTTKPAPT	377
Db	320	EPAPTTPKEPAPTITTKSAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPTPTTKPAPT	379
QY	378	TKEPAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPAPTTPKEPTPTTKPAPTTPKSA	437
Db	380	TKEPAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPAPTTPKEPTPTTKPAPTTPKSA	439
QY	438	PTTKPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPTPTTKPAPTTPK	497
Db	440	PTTKPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPTPTTKPAPTTPK	499
QY	498	EPAPTTPKPAPTTPKEPAPTTPKPAPTTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPT	557
Db	500	EPAPTTPKPAPTTPKEPAPTTPKPAPTTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPT	559
QY	558	TTPEBPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK	617
Db	560	TTPEBPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK	619
QY	618	GTAPTTLKEPAPTTPKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	677
Db	620	GTAPTTLKEPAPTTPKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	679
QY	678	KEPAPTTPKPAKPAKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	737
Db	680	KEPAPTTPKPAKPAKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	739
QY	738	PKPAPTTPKPAKPAKPAKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	797
Db	740	PKPAPTTPKPAKPAKPAKPAKPAKELAPTTTKPTSTTSKPAPTTPKGTAPTTPKEPAPTTP	799
QY	798	NSPKPAPVPTTKPAATKPEMTTAKDKTERDLTTTPTTAAKPMKETAATTTTEKTE	857
Db	800	NSPKPAPVPTTKPAATKPEMTTAKDKTERDLTTTPTTAAKPMKETAATTTTEKTE	859
QY	858	SKITATTTQVSTTTQDTPPKITTLTKTTTLAPKVTITTKITITTEIMNKPEETAKPKDR	917
Db	860	SKITATTTQVSTTTQDTPPKITTLTKTTTLAPKVTITTKITITTEIMNKPEETAKPKDR	919
QY	918	ATNSKATTPKQKPKAPKPTSTKPKTMRVRKPKTTPPKMTSTMPELNPTSTRAE	977
Db	920	ATNSKATTPKQKPKAPKPTSTKPKTMRVRKPKTTPPKMTSTMPELNPTSTRAE	979

QY	978	AWLOTTTPNQTPNSKLVEVNPKESEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVN	1037
Db	980	AWLOTTTPNQTPNSKLVEVNPKESEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVN	1039
QY	1038	QGIILNPMLSDETNI CNCKPVDGLTTLRNGTLVAFRGHYFWMLSFFSPSPARRITVWG	1097
Db	1040	QGIILNPMLSDETNI CNCKPVDGLTTLRNGTLVAFRGHYFWMLSFFSPSPARRITVWG	1099
QY	1098	IPSPIDTVTRCNCCEGKTFEFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTA	1157
Db	1100	IPSPIDTVTRCNCCEGKTFEFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTA	1159
QY	1158	KYKNWPESVVFYFKGSGSIQYIYKQBPVQKCPGRRPALNVPVYGMTQVRRRRRERAIGP	1217
Db	1160	KYKNWPESVVFYFKGSGSIQYIYKQBPVQKCPGRRPALNVPVYGMTQVRRRRRERAIGP	1219
QY	1218	SOHTTIRIQYSPARLAYODKGVHLNVEKVSILNRGLPNVVTSAISLNPINRKPQDYDYAF	1277
Db	1220	SOHTTIRIQYSPARLAYODKGVHLNVEKVSILNRGLPNVVTSAISLNPINRKPQDYDYAF	1279
QY	1278	SKDOYNNIDVPSRTARAITTRSGQTLSKWNCP	1311
Db	1280	SKDOYNNIDVPSRTARAITTRSGQTLSKWNCP	1313
RESULT 11			
US-10-124-557-46			
; Sequence 46, Application US/10124557			
; Publication No. US20020137894A1			
GENERAL INFORMATION:			
APPLICANT: Turner, Katherine			
Clark, Stephen C.			
Jacobs, Kenneth			
Hewick, Rodney M.			
Gesner, Thomas G.			
TITLE OF INVENTION: Megakaryocyte Stimulating Factors			
NUMBER OF SEQUENCES: 143			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Genetics Institute, Inc.			
STREET: 87 CambridgePark Drive			
CITY: Cambridge			
STATE: Massachusetts			
COUNTRY: U.S.A.			
ZIP: 02140			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patentin Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/10/124,557			
FILING DATE: 16-Apr-2002			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 07/643,502			
FILING DATE: 18-JAN-1991			
APPLICATION NUMBER: US 07/546,114			
FILING DATE: 29-JUN-1990			
APPLICATION NUMBER: US 07/457,196			
FILING DATE: 29-DEC-1989			
APPLICATION NUMBER: US 07/390,901			
FILING DATE: 08-AUG-1989			
ATTORNEY/AGENT INFORMATION:			
NAME: Cserr, Luann			
REGISTRATION NUMBER: 31,822			
REFERENCE/DOCKET NUMBER: GI 5190			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (617)876-1170			
TELEFAX: (617)876-5851			
INFORMATION FOR SEQ ID NO: 46:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 1320 amino acids			
TYPE: amino acid			

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 96.0%; Score 6784.9; DB 13; Length 1320;
Best Local Similarity 93.3%; Pred. No. 1.8e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIVLLLLSVFVITQVSSDLSSCAGRCGEGSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIVLLLLSVFVITQVSSQ----- 25
QY 61 KRVTAEALCKGRCFSPERGERECDDAOCKYDKCCPDYSEFCA----- 105
DB 26 -----ELSCGKRCFSPERGERECDDAOCKYDKCCPDYSEFCAEVHNPPTSPSSKKAP 79
QY 106 -----EVKDNKKKRTKKKTPKPPVVVDEAG 130
DB 80 PPSGASQTIKSTTKRSPKPNKKTKVIESEBIEVKNKKRTKKKTPKPPVVVDEAG 139
QY 131 SGLDNGDFKVTPTDSTTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETT 190
DB 140 SGLDNGDFKVTPTDSTTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETT 199
QY 191 VETKETTTNNKQTSIDGKEKTSKAKTQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 250
DB 200 VETKETTTNNKQTSIDGKEKTSKAKTQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 259
QY 251 LITPKETPTTPKEPASTTKEPTPTTKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKE 310
DB 260 LITPKETPTTPKEPASTTKEPTPTTKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKE 319
QY 311 PAPTTPKEPAPTTTKPEPAPTTKSAPTTKPEPAPTTPKPEPAPTTPKPEPAPTTT 370
DB 320 PAPTTPKEPAPTTTKPEPAPTTKSAPTTKPEPAPTTPKPEPAPTTPKPEPAPTTT 379
QY 371 PKPEPAPTTKEPAPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 430
DB 380 PKPEPAPTTKEPAPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 439
QY 431 PTTTKSAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKE 490
DB 440 PTTTKSAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKE 499
QY 491 PAPTTPKEPAPTTTKPEPAPTTAPKEPAPTTPKETAPTTPKLTPPTPKLAPTTPEKAPT 550
DB 500 PAPTTPKEPAPTTTKPEPAPTTAPKEPAPTTPKETAPTTPKLTPPTPKLAPTTPEKAPT 559
QY 551 TPESLAPTTPEPTPTTPEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTTKE 610
DB 560 TPESLAPTTPEPTPTTPEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTTKE 619
QY 611 TAPTTPKGTAPTTLKPEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTTPKGTAPTTPK 670
DB 620 TAPTTPKGTAPTTLKPEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTTPKGTAPTTPK 679
QY 671 EPAPTTPKEPAPTTPKGTAPTTLKPEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTT 730
DB 680 EPAPTTPKEPAPTTPKGTAPTTLKPEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTT 739
QY 731 KETAPTTKPEPAPTTPKKAPAPTTTPEPTTSEVSTPTTKEPTTIHKSDESTPELSAE 790
DB 740 KETAPTTKPEPAPTTPKKAPAPTTTPEPTTSEVSTPTTKEPTTIHKSDESTPELSAE 799
QY 791 PTKALENSPKGPGVPTTKTAAKPEMTTAKDKITRDLRTPETTTAAAPKMTKETAT 850
DB 800 PTKALENSPKGPGVPTTKTAAKPEMTTAKDKITRDLRTPETTTAAAPKMTKETAT 859
QY 851 TTEKTTESKITAATTQVSTTTTODTTPPKITTLKTTTLAPKVTTTKITTEINMKPEE 910
DB 860 TTEKTTESKITAATTQVSTTTTODTTPPKITTLKTTTLAPKVTTTKITTEINMKPEE 919

QY 911 TAKPKDRATNSKATTPKQKTPKAPKXPTSTTKKPKTMRVRKPKTTPTRKMTSTMPELN 970
DB 920 TAKPKDRATNSKATTPKQKTPKAPKXPTSTTKKPKTMRVRKPKTTPTRKMTSTMPELN 979
QY 971 PTSRIAEAMLQTTTRPNQTPNSKLVENPKSEADAGAGETPHMLLRPHVMPETPDMD 1030
DB 980 PTSRIAEAMLQTTTRPNQTPNSKLVENPKSEADAGAGETPHMLLRPHVMPETPDMD 1039
QY 1031 YLPVRPNQIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSPAR 1090
DB 1040 YLPVRPNQIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSPAR 1099
QY 1091 RITEVWGIPIPIVFTTRCNCCEKTFKFKOSQYWRFTNDIKDAGPKPIFKFGGLTGOI 1150
DB 1100 RITEVWGIPIPIVFTTRCNCCEKTFKFKOSQYWRFTNDIKDAGPKPIFKFGGLTGOI 1159
QY 1151 VAALSTAKYKNWPESVYFFKRGSGSQOYLYKQEPVQKCPGRPALNPVYVGMETQVRRR 1210
DB 1160 VAALSTAKYKNWPESVYFFKRGSGSQOYLYKQEPVQKCPGRPALNPVYVGMETQVRRR 1219
QY 1211 FERRAIGPSOITHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1270
DB 1220 FERRAIGPSOITHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1279
QY 1271 GYDYAFSKDQYNNIDVPSRTARAITRRSGQTLKQVWYNCP 1311
DB 1280 GYDYAFSKDQYNNIDVPSRTARAITRRSGQTLKQVWYNCP 1320

RESULT 12
US-10-124-557-60
; Sequence 60, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseriz, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 96.0%; Score 6784.9; DB 13; Length 1320;
Best Local Similarity 93.3%; Pred. No. 1.8e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNDVNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFVIQVSSQ-----25
QY 61 KRVCATLSKGRCFESFERGREGCDCAQCKYDKCCPDYESFCA-----105
DB 26 -----ELSKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
QY 106 -----EYKDNKKNRTKKKPTPKPPVVDGAG 130
DB 80 PPSGASQTIKTSRSPKPNKKTKVIESEBITEVKDNKKNRTKKKPTPKPPVVDGAG 139
QY 131 SGLDNGDFKVTDTSTTQHNKYSTSPKITTAKPINRPSLPNPSDTSKETSITVNKETT 190
DB 140 SGLDNGDFKVTDTSTTQHNKYSTSPKITTAKPINRPSLPNPSDTSKETSITVNKETT 199
QY 191 VETKETTNNKQSTDGKCKTSAKETQSTKTSKADLAPTSKVLAKEPTTKAETTTKGA 250
DB 200 VETKETTNNKQSTDGKCKTSAKETQSTKTSKADLAPTSKVLAKEPTTKAETTTKGA 259
QY 251 LTPKKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKEPAPTTPKE 310
DB 260 LTPKKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKEPAPTTPKE 319
QY 311 PAPTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 370
DB 320 PAPTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 379
QY 371 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEA 430
DB 380 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEA 439
QY 431 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 490
DB 440 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 499
QY 491 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 550
DB 500 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 559
QY 551 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 610
DB 560 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 619
QY 611 TAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 670
DB 620 TAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 679
QY 671 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 730
DB 680 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 739
QY 731 KETAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 790
DB 740 KETAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 799
QY 791 PTPKALENSPKEPGVPTTKTTPAATKPEMTTTAKDKTTERDITTTETTTAAPKMTKETAT 850
DB 800 PTPKALENSPKEPGVPTTKTTPAATKPEMTTTAKDKTTERDITTTETTTAAPKMTKETAT 859

QY 851 TTEKTESKITATTTQVTSITTTQDTPPKITLTLTKTLTTLAPKVTTTKITTTTEIMNKPEE 910
DB TTEKTESKITATTTQVTSITTTQDTPPKITLTLTKTLTTLAPKVTTTKITTTTEIMNKPEE 919
QY 911 TAKPKDRATNSKATTPKPKOKETKAPKPTSTKTKTMTPRVRKPKTTPPRKWTSTMPBLN 970
DB TAKPKDRATNSKATTPKPKOKETKAPKPTSTKTKTMTPRVRKPKTTPPRKWTSTMPBLN 979
QY 971 PTSRIAEAMLQTTTRPNQNTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPMD 1030
DB PTSRIAEAMLQTTTRPNQNTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPMD 1039
QY 1031 YLPRVFNQGIILNMLSDENNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1090
DB YLPRVFNQGIILNMLSDENNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1099
QY 1091 RITEVWGIPLSPIDVTRCNCEGKTFEKKDQSOYWRFTNDIKDAGYKPIFKGFGGLTQOI 1150
DB RITEVWGIPLSPIDVTRCNCEGKTFEKKDQSOYWRFTNDIKDAGYKPIFKGFGGLTQOI 1159
QY 1151 VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVQKCPGRPALNYPVYVGMTOVRRR 1210
DB VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVQKCPGRPALNYPVYVGMTOVRRR 1219
QY 1211 FERAIGPSQHTTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1270
DB FERAIGPSQHTTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKP 1279
QY 1271 GYDYAFSKDQYNNIDVPSRTARITTRSGOTLSKQWYNCP 1311
DB GYDYAFSKDQYNNIDVPSRTARITTRSGOTLSKQWYNCP 1320

RESULT 13
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 96.0%; Score 6780.6; DB 13; Length 1363;
Best Local Similarity 90.5%; Pred. No. 2.3e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSACAGCGEGYSDATCNCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
QY 61 KRVCCTAELSCGRCFBSFERGECDCDAOCKKYDKCCPDYESFCAE----- 106
DB 26 -----ELSCGRCFBSFERGECDCDAOCKKYDKCCPDYESFCAEYHNPTSPPSKKAP 79
QY 107 ----- 106
DB 80 PPSGASQTKSTTKRSPKPPNKKTKVIESEBITEHSVSENQESSSSSSSSSTIW 139
QY 107 -----VKDNKQRTKKPKPPVVDVDEAGSLDNGDPKVTTPDTST 147
DB 140 KIKSSKNSAANRELQKKLVKDNKNTKKPKPPVVDVDEAGSLDNGDPKVTTPDTST 199
QY 148 TQNKVYSTPKITAKINRPSLPNSDTSKETSIVNKETTVETKETTNNKQSTDG 207
DB 200 TQNKVYSTPKITAKINRPSLPNSDTSKETSIVNKETTVETKETTNNKQSTDG 259
QY 208 KEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 267
DB 260 KEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 319
QY 268 TTPKEPTTPKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
DB 320 TTPKEPTTPKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
QY 328 APPTTKSAPTTPKPEAPTTPKKPAPTTPKPEAPTTPKPEPTTPPKPEAPTTKPEAPTTPK 387
DB 380 APPTTKSAPTTPKPEAPTTPKKPAPTTPKPEAPTTPKPEPTTPPKPEAPTTKPEAPTTPK 439
QY 388 EPAPTAPKPAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTPKPEAPTT 447
DB 440 EPAPTAPKPAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTPKPEAPTT 499
QY 448 TTKSAPTTPKPEPTTKPEAPTTPKKPAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKKPK 507
DB 500 TTKSAPTTPKPEPTTKPEAPTTPKKPAPTTPKPEAPTTPKKPAPTTPKPEAPTTTKKPK 559
QY 508 APAPTAPKPEAPTTPKKPAPTTPPKKPAPTTPPKKPAPTTPPKKPAPTTPPKKPAPTTPPK 567
DB 560 APAPTAPKPEAPTTPKKPAPTTPPKKPAPTTPPKKPAPTTPPKKPAPTTPPKKPAPTTPPK 619
QY 568 PEETAPPTPKAAANTPKPEAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKGTAPTTLKPE 627
DB 620 PEETAPPTPKAAANTPKPEAPTTTKPEAPTTTKPEAPTTTKGTAPTTLKPE 679
QY 628 APPTPKPEAPKELAPTTTKPEPTSTSDKPAPTTPKGTAPTTKPEAPTTPKPEAPTTPKG 687
DB 680 APPTPKPEAPKELAPTTTKPEPTSTSDKPAPTTPKGTAPTTKPEAPTTPKG 739
QY 688 TAPPTLKPEAPPTPKKPAKELAPTTTKGPTSTSDKPAPTTPKETAPTTKPEAPTTPK 747
DB 740 TAPPTLKPEAPPTPKKPAKELAPTTTKGPTSTSDKPAPTTPKETAPTTKPEAPTTPK 799

748 KPAPTTPEPTTSEVSTPTTTKEPTTHKSPDESTPELSAEPPTKALENSPKPEGVPT 807
DB KPAPTTPEPTTSEVSTPTTTKEPTTHKSPDESTPELSAEPPTKALENSPKPEGVPT 859
QY TKTPAATKPEMTTAKDKTTERDLATTTETTTAAAPKMKETATTTETTESKITATTTOV 867
DB TKTPAATKPEMTTAKDKTTERDLATTTETTTAAAPKMKETATTTETTESKITATTTOV 919
QY TSSTTTQDTPPKITTLTKTLAPKVTTKTITTEIMNKPETAKPKDRATNSKATTPK 927
DB TSSTTTQDTPPKITTLTKTLAPKVTTKTITTEIMNKPETAKPKDRATNSKATTPK 979
QY POKPKAPKPKPTSTKKPKTMRVRKPKTTPTPRKMSTMPELNPTSRIAEAMLQTTTRN 987
DB POKPKAPKPKPTSTKKPKTMRVRKPKTTPTPRKMSTMPELNPTSRIAEAMLQTTTRN 1039
QY QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGGIIINPMLS 1047
DB QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGGIIINPMLS 1099
QY DETNINCNGPVDGLTTLRNGTLVAPRGHYFWMLSPEPSPAPRRI TEVWGIPSPIDTFT 1107
DB DETNINCNGPVDGLTTLRNGTLVAPRGHYFWMLSPEPSPAPRRI TEVWGIPSPIDTFT 1159
QY RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAALSTAKYKNWPESVY 1167
DB RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAALSTAKYKNWPESVY 1219
QY FFKRGSGIOQYIYKQBPVQKCGRRPALNYPVVGEMTOVRRRFFERAIGPSQHTIRIQY 1227
DB FFKRGSGIOQYIYKQBPVQKCGRRPALNYPVVGEMTOVRRRFFERAIGPSQHTIRIQY 1279
QY SPARLAYQDGVLUHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYYNIDV 1287
DB SPARLAYQDGVLUHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYYNIDV 1339
QY PSRTARAITRSQGLTSKVWYNCP 1311
DB PSRTARAITRSQGLTSKVWYNCP 1363

RESULT 14
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114


```

; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match          79.3%; Score 5603.7; DB 13; Length 1140;
Best Local Similarity 91.8%; Pred No. 6.4e-128;
Matches 1047; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVTAELSCKGRCFSEFREGREDCDCAQCKYDKCDDYSEFCAE-----106
DB 61 KRVTAELSCKGRCFSEFREGREDCDCAQCKYDKCDDYSEFCAEVHNPTSPSSKKAP 120
QY 107 -----106
DB 121 PPSGASQTIKSTKRSKPPNKKTKKVISEBITEHSVSENQESSSSSSSSSTIW 180
QY 107 -----VKNNKKNRTKKKPPKPPVVDVDEAGSGLDNGDFKVTTPDTS 147
DB 181 KIKSSKNSAANRELQKKLVKONKKNRTKKKPPKPPVVDVDEAGSGLDNGDFKVTTPDTS 240
QY 148 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKITTNNKQSTDG 207
DB 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKITTNNKQSTDG 300
QY 208 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 267
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPE 327
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPE 420
QY 328 APTTTKSAPTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 387
DB 421 APTTTKSAPTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 480
QY 388 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 447
DB 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 540
QY 448 TTKSAPTTPKSPSTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 507
DB 541 TTKSAPTTPKSPSTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 600
QY 508 APTAPKEPAPTTPKETAPTTPKLTTPPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 567
DB 601 APTAPKEPAPTTPKETAPTTPKLTTPPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 660
QY 568 PEEAPPTPKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 627
DB 661 PEEAPPTPKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 720

```

```

QY 628 APTTPKPKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTPKBPAPTTPKBPAPTT 687
DB 721 APTTPKPKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTPKBPAPTTPKBPAPTT 780
QY 688 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTPKBPAPTT 747
DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTPKBPAPTT 840
QY 748 KPAPTTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPE 807
DB 841 KPAPTTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPE 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPEITTAAPKMKETATTTTEKTTESKITATT 867
DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPEITTAAPKMKETATTTTEKTTESKITATT 960
QY 868 TSTTTQDTPFKITLTKITTLAPKVTTTKITTTTEIMNKPEETAKPKDRATNSKATPK 927
DB 961 TSTTTQDTPFKITLTKITTLAPKVTTTKITTTTEIMNKPEETAKPKDRATNSKATPK 1020
QY 928 POKPTKAPKKPTSTKPKTMPRVRKPKTTTPRKMTSTMPELNPTSRISAEAMLOTTTRPN 987
DB 1021 POKPTKAPKKPTSTKPKTMPRVRKPKTTTPRKMTSTMPELNPTSRISAEAMLOTTTRPN 1080
QY 988 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1047
DB 1081 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140

```

RESULT 15

US-10-124-557-74

; Sequence 74, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)876-1170

; TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 77.9%; Score 5500.7; DB 13; Length 1038;
Best Local Similarity 79.2%; Pred. No. 1.8e-125;
Matches 1038; Conservative 0; Mismatches 0; Indels 273; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGGYSDATCNDYNCQHMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ-----25

QY 61 KRVTAEISCKGRCHESFERGECDDAOCKYDKCCPDYSEFCAEVKDNKNRTKKKPT 120
DB 26 -----ELSKGRCHESFERGECDDAOCKYDKCCPDYSEFCAEVKDNKNRTKKKPT 79

QY 121 PKPPVVDEAGSLDNGDFKVTPTDTSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKE 180
DB 80 PKPPVVDEAGSLDNGDFKVTPTDTSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKE 139

QY 181 TSLTVNKETTVETKETTNTNKTSTDGKEKTTSAKETQSIKTSKDLAPTSKVLAKPTP 240
DB 140 TSLTVNKETTVETKETTNTNKTSTDGKEKTTSAKETQSIKTSKDLAPTSKVLAKPTP 199

QY 241 KAEITTKGPAITTKPEPTTPKEPASTTPEKPTTIKSAPTTKPEPAPTTTKSAPTTP 300
DB 200 KAEITTKGPAITTKPEPTTPKEPASTTPEKPTTIKSAPTTKPEPAPTTTKSAPTTP 259

QY 301 KEPAITTKPEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAP 360
DB 260 KEPAITTKPEPAPTTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAP 319

QY 361 TTPKEPTTPTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKE 420
DB 320 TTPKEPTTPTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKE 379

QY 421 PSPTTKPEPAPTTTKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTT 480
DB 380 PSPTTKPEPAPTTTKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTT 439

QY 481 KKDAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 540
DB 440 KKDAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 499

QY 541 PTTPEKPAITPEELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 600
DB 500 PTTPEKPAITPEELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 559

QY 601 KEPAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 660
DB 560 KEPAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 619

QY 661 PKGTAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 720
DB 620 PKGTAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 679

QY 721 TSDKPAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 780
DB 680 TSDKPAITTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 739

QY 781 DESTPELSAETPPKALENSKPEPGVPTTKTAAKPEMTTTAKOKTTERDLRTTPEITTA 840
DB 740 DESTPELSAETPPKALENSKPEPGVPTTKTAAKPEMTTTAKOKTTERDLRTTPEITTA 799

QY 841 APKMTKETATTTTEKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTITTKKIT 900
DB 800 APKMTKETATTTTEKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTITTKKIT 859

Search completed: October 13, 2004, 11:53:32
Job time : 142.379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 144.144 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKTPDYLLLLSVFVIQ.....ARAITRRSGQTLSKVWVNC P 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 segs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_02:*

1: uniprot_prot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7044.7	99.7	1404	2	Q92954	Q92954 homo sapien
2	7029.7	99.5	1404	2	Q9BX49	Q9BX49 homo sapien
3	4370.6	61.9	933	2	Q6ZM25	Q6ZM25 homo sapien
4	4370.6	61.9	933	2	BAD18580	BAD18580 h cdna fl
5	3827.7	54.2	1054	2	Q9JM99	Q9JM99 mus musculus
6	1714.7	24.3	401	2	Q77765	Q77765 bos taurus
7	1398.2	19.8	5179	1	MUC2_HUMAN	MUC2_HUMAN
8	1338	18.9	1225	2	Q9VR49	Q9VR49
9	1333	18.9	251	2	Q6DNC4	Q6DNC4
10	1311.6	18.6	1761	2	Q7KTF6	Q7KTF6
11	1311.6	18.6	1761	2	AA564673	AA564673
12	1271.2	18.0	3150	2	Q7PMD5	Q7PMD5
13	1211.6	17.2	3409	2	Q6SSE6	Q6SSE6
14	1211.6	17.2	3409	2	AA507044	AA507044
15	1183	16.7	251	2	Q6QJF3	Q6QJF3
16	1183	16.7	251	2	AA583977	AA583977
17	1154.2	16.3	1664	1	SLP1_CLOTH	SLP1_CLOTH
18	1150.6	16.3	3889	2	Q6SSE8	Q6SSE8
19	1150.6	16.3	3889	2	AA507042	AA507042
20	1112.2	15.7	5703	1	MUSB_HUMAN	MUSB_HUMAN
21	1102.5	15.6	1349	2	Q8WQ4	Q8WQ4
22	1102	15.6	3432	2	Q8TR51	Q8TR51
23	1102	15.6	3458	2	Q8TR52	Q8TR52
24	1096.4	15.5	1795	2	Q76894	Q76894
25	1093.3	15.5	1079	2	Q9N4S7	Q9N4S7
26	1084.6	15.4	9234	2	Q7KTP5	Q7KTP5
27	1084.6	15.4	9234	2	AA10531	AA10531
28	1079	15.3	23015	2	Q8IQ18	Q8IQ18
29	1079	15.3	23015	2	AA10358	AA10358
30	1061.7	15.0	34350	2	Q8WZ42	Q8WZ42
31	1054.5	14.9	2284	2	Q9VPG1	Q9VPG1

32	1038	14.7	7962	2	Q10465	Q10465 homo sapien
33	1034.5	14.6	972	2	Q7QKK7	Q7QKK7 anopheles g
34	1034.2	14.6	1607	2	Q9H6Q5	Q9H6Q5 phytophthor
35	1034.2	14.6	1607	2	AA74661	AA74661 phytophthor
36	1021.4	14.5	2187	2	P70670	P70670 mus musculus
37	1011.3	14.3	1489	2	Q96449	Q96449 phytophthor
38	1010.5	14.3	1274	2	Q20007	Q20007 caenorhabdi
39	1001.8	14.2	1458	2	Q757N5	Q757N5 ashbya goss
40	1001.8	14.2	1458	2	AA52662	AA52662 ashbya goss
41	1001.8	14.2	4498	2	Q9W223	Q9W223 drosophila
42	983.6	13.9	10578	2	Q8ISF5	Q8ISF5 caenorhabdi
43	983.2	13.9	1480	2	Q9LIE8	Q9LIE8 arabidopsis
44	981.6	13.9	18519	2	Q8ISF6	Q8ISF6 caenorhabdi
45	981.6	13.9	18534	2	Q8ISF7	Q8ISF7 caenorhabdi

ALIGNMENTS

RESULT 1

Q92954
ID Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Megakaryocyte stimulating factor.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity.";
RN Blood 78:279-279(1991).
RL [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier
RL Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U70136; AAB09089.1; -;
DR HSSP; P04004; 10C0.
DR Genew; HGNC:9364; PKG4.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;


```
QY 107 -----VKDNKNRTKKXPTPKPPVVDGAGSLDNGDFKVTPTDST 147
Db 181 KIKSSKNSAANREIQKKLVKDNKNRTKKXPTPKPPVVDGAGSLDNGDFKVTPTDST 240
QY 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 207
Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
QY 268 TTPKEPTPTTIKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPK 327
Db 361 TTPKEPTPTTIKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPK 420
QY 328 APPTTKSAPTTTPKEPAPTTPKAPPTPKGAPPTTPKEPTTTPKSAPTTTPKEPAPTTPK 387
Db 421 APPTTKSAPTTTPKEPAPTTPKAPPTPKGAPPTTPKEPTTTPKSAPTTTPKEPAPTTPK 480
QY 388 EPAPTAPKAPPTTTPKEPAPTTPKEPAPTTPKBSPTTPKEPAPTTPKSAPTTTPKEPAPT 447
Db 481 EPAPTAPKAPPTTTPKEPAPTTPKBSPTTPKEPAPTTPKBSPTTPKEPAPTTPKSAPTT 540
QY 448 TTKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPKAPPTTPKSAPTTTPKEPAPTTPK 507
Db 541 TTKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTPKAPPTTPKSAPTTTPKEPAPTTPK 600
QY 508 APPTAPKEPAPTTPKEPAPTTPKKLTPPTPKLAPPTTPKEPAPTTPKSAPTTTPKEPAPT 567
Db 601 APPTAPKEPAPTTPKEPAPTTPKKLTPPTPKLAPPTTPKEPAPTTPKSAPTTTPKEPAPT 660
QY 568 PEPAAPTTPKAAAPNTPKAPPTTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPK 627
Db 661 PEPAAPTTPKAAAPNTPKAPPTTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPK 720
QY 628 APPTPKAPKAPKELAPPTTTPKEPTSTTSKAPPTTPKAPPTTPKAPPTTPKSAPTTTPK 687
Db 721 APPTPKAPKAPKELAPPTTTPKEPTSTTSKAPPTTPKAPPTTPKAPPTTPKSAPTTTPK 780
QY 688 TAPPTLKEPAPTTPKAPKAPKELAPPTTTPKGTSTTSKAPPTTPKAPPTTPKSAPTTTPK 747
Db 781 TAPPTLKEPAPTTPKAPKAPKELAPPTTTPKGTSTTSKAPPTTPKAPPTTPKSAPTTTPK 840
QY 748 KPAPTTPPTTSPVSTPTTTPKEPTTHKSDSTPESAPPTPKALENSKEPAGVPT 807
Db 841 KPAPTTPPTTSPVSTPTTTPKEPTTHKSDSTPESAPPTPKALENSKEPAGVPT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPETTTTAAAPKWKETATTTTETTESKITATTQV 867
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTTAAAPKWKETATTTTETTESKITATTQV 960
QY 868 TSTTTQDTPPKKITLKITTLAPKVTTTKKTTTTHIMNKBETAKPKDRATNSKATTPK 927
Db 961 TSTTTQDTPPKKITLKITTLAPKVTTTKKTTTTHIMNKBETAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKAPKSTTKPKTPMPVRPKPTPTPKMTSTMPELNPTSRIAEAMLOTTTRN 987
Db 1021 POKPTKAPKAPKSTTKPKTPMPVRPKPTPTPKMTSTMPELNPTSRIAEAMLOTTTRN 1080
QY 988 QTPNSKLVENPKSBDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1047
Db 1081 QTPNSKLVENPKSBDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140
QY 1048 DETNTCNGKPVUGLTLTRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGISPIDTFT 1107
Db 1141 DETNTCNGKPVUGLTLTRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGISPIDTFT 1200
QY 1108 RCNCEGKTFFFKDSQVWRFTNDIKDAGYKPKTFKFGGLTGQIVAAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTFFFKDSQVWRFTNDIKDAGYKPKTFKFGGLTGQIVAAALSTAKYKNWPESVY 1260
```

```
QY 1168 FFKGGSIQIYYIKQEPVQKCPGRRPALNVPVYGMTQVRRRRRPERAIGPSQTHIRIQY 1227
Db 1261 FFKGGSIQIYYIKQEPVQKCPGRRPALNVPVYGMTQVRRRRRPERAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSTLWEGLPNVVTSALSLNIRKPDGYDYAFSKDQYINIDV 1287
Db 1321 SPARLAYQDKGVLHNEVKVSTLWEGLPNVVTSALSLNIRKPDGYDYAFSKDQYINIDV 1380
QY 1288 PSRTARAITTRSGQTLSSKVVWVNC 1311
Db 1381 PSRTARAITTRSGQTLSSKVVWVNC 1404

RESULT 3
Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBL_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yaenashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131434; BAD18580.1; ...
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF00045; Hemopexin; 2.
DR Pfam: PF01033; Somatomedin_B; 1.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 1.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;
```

Query Match 61.9%; Score 4370.6; DB 2; Length 933;
Best Local Similarity 59.6%; Pred. No. 1e-83;
Matches 839; Conservative 0; Mismatches 1; Indels 564; Gaps 3;

```
QY 1 MAWKTLPIYLLLLLSVFVIQVSSODLSSCAGCGGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFVIQVSSO----- 25
QY 61 KRVTAEILCKGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAE----- 106
Db 26 -----ELSCGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79
QY 107 ----- 106
Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEHSVSENQSSSSSSSSSSTIR 139
QY 107 -----VKDNKNRTKKXPTPKPPVVDGAGSLDNGDFKVTPTDST 147
Db 140 KIKSSKNSAANREIQKKLVKDNKNRTKKXPTPKPPVVDGAGSLDNGDFKVTPTDST 199
QY 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 207
Db 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 259
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 267
```

Db 260 KETTSAKETQSIKTSKDLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTTKEPAS 319
QY 268 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEP 327
Db 320 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKE 370
QY 328 APTTTKSAPTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKEPTTTPKEPAPTTTKEPAPTTPK 387
Db 371 ----- 370
QY 388 EPAPTAPKKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPT 447
Db 371 ----- 370
QY 448 TTKSAPTTPKERSPTTTPKEPAPTTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKEPAPTTTXXP 507
Db 371 ----- 370
QY 508 APTAPKEPAPTTTPKETAPTTPPKLTPTTPEKLAPTTPKEPAPTTTPPELAPTTPEEPTPTT 567
Db 371 ----- 370
QY 568 PPEPAPTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP 627
Db 371 ----- 370
QY 628 APTTPKKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEPAPTTTPKEPAPTTTPKPG 687
Db 371 ----- 370
QY 688 TAPTTLKEPAPTTTPKBPAPKELAPTTTKEPTSTSDKAPATTPKETAPTTPKEPAPTTTPK 747
Db 371 ----- 370
QY 748 KPAPTTPPETPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSKPEGPVT 807
Db 371 -PAPTTPPETPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSKPEGPVT 429
QY 808 TKTPAATKPEMTTAKDKTTERDLRTPTTTPAAPKMTKETATTTTEKTESKITATTQV 867
Db 430 TKTPAATKPEMTTAKDKTTERDLRTPTTTPAAPKMTKETATTTTEKTESKITATTQV 489
QY 868 TSTTTQDTPFKITTLKTTLAPKVTITTKITTIIMNKPEETAKPKDRATNSKATTPK 927
Db 490 TSTTTQDTPFKITTLKTTLAPKVTITTKITTIIMNKPEETAKPKDRATNSKATTPK 549
QY 928 POKETKAPKPTSTPKPKTMVRVKPTTPTTPKONTSTMPELNPTSRIAEAMLQTTTRPN 987
Db 550 POKETKAPKPTSTPKPKTMVRVKPTTPTTPKONTSTMPELNPTSRIAEAMLQTTTRPN 609
QY 988 QTNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1047
Db 610 QTNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 669
QY 1048 DETNI CNKGPVDGLTTLNGLTLVAFRGHYFWMLSPEFSPSPSARITTEVWGTPSIDIVFT 1107
Db 670 DETNICNGKPDVGLTTLNGLTLVAFRGHYFWMLSPEFSPSPSARITTEVWGTPSIDIVFT 729
QY 1108 RCNCEGHTTFFFKDSQYWRFNNDIKDAGYKPKIFKFGGLTQCIIVAALSTAKYKWPESVY 1167
Db 730 RCNCEGHTTFFFKDSQYWRFNNDIKDAGYKPKIFKFGGLTQCIIVAALSTAKYKWPESVY 789
QY 1168 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGEMTQVRRRRFERAIGPSQTHIRIQY 1227
Db 790 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGEMTQVRRRRFERAIGPSQTHIRIQY 849
QY 1228 SPARLAYODKGLVHNEVKVSTLWGLNPNVTSALSLPNIRKPDGDYVYAFSKOQYXNIDV 1287
Db 850 SPARLAYODKGLVHNEVKVSTLWGLNPNVTSALSLPNIRKPDGDYVYAFSKOQYXNIDV 909
QY 1288 PSRTARAITTRSGQTLKSVWYNCP 1311

Db 910 PSRTARAITTRSGQTLKSVWYNCP 933

RESULT 4
ID BAD18580 PRELIMINARY; PRT; 933 AA.
AC BAD18580;
DT 12-MAY-2004 (TReMBLrel. 27, Created)
DT 12-MAY-2004 (TReMBLrel. 27, Last sequence update)
DE CDNA FLJ16561 fis, clone SYN04003981, moderately similar to Homo sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular superficial zone protein, camptodactyly, arthropathy, coxa vara, pericarditis syndrome) (PRG4).
DE Homo sapiens (human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Synovial membrane tissue;
RC Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuko Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1; -;
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 61.9%; Score 4370.6; DB 2; Length 933;
Best Local Similarity 59.8%; Pred. No. 1e-83; 1; Indels 564; Gaps 3;
Matches 839; Conservative 0; Mismatches 1;
QY 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQ ----- 25
QY 61 KRVCYTAELSCGRCFSPSFERGRCDCDAQCKKYDKCCPDYESFCAE ----- 106
Db 26 -----ELSCGRCFSPSFERGRCDCDAQCKKYDKCCPDYESFCAEVHNFTSPSSKKAP 79
QY 107 ----- 106
Db 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEIETEHVSSENOESSSSSSSSSTIR 139
QY 107 -----VKDNKNRTKKKTPPKPPVVDVDEAGSLDNGDFKVTTPDTST 147
Db 140 KIKSSKNSAANRELQKKLVKDNKNKNTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
QY 148 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKTTTNTKQSTDG 207
Db 200 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKTTTNTKQSTDG 259
QY 208 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTPKEPAS 267
Db 260 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTPKEPAS 319
QY 268 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEP 327
Db 320 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKE 370
QY 328 APTTTKSAPTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKEPTTTPKEPAPTTTKEPAPTTPK 387
Db 371 ----- 370
QY 388 EPAPTAPKKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPAPT 447
Db 371 ----- 370


```

QY 623 TLKEPAPPTPKKAPKELAPTTTKBPTSTSDKAPPTPKGTAPTTKBPAPTTKBPAP 682
DB 539 --KEPEPTPKKP-----EPTTPKEPVPTTKPEBP 567
QY 683 TTPKGTAPTTKBPAPTTKPKAPKELAPTTTKGPTSTSDKAPPTTKETAPTTKPEPA 742
DB 568 TTP-----KEPEPTPKPEP-----TTRKEPEPTTKPEPTPKPEP 606
QY 743 PTTPKKAPPTTPEPTTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETPKALENSPKE 802
DB 607 PTPKKEPTT----- 617
QY 803 PGVPTTKTAPATKPEMTTAKOKTTERDLRTTPTTAAKPMKTKETATITKTESKITA 862
DB 618 -----TSPKT----- 622
QY 863 TTTQVTSITQDTPPKITLTKTTLAPKVTTTKTITTEIMNKPEETAAPKPDATNSK 922
DB 623 -----TTLKATLAPKVTAPE-----EIQNKPEETTPASEDSDDSK 659
QY 923 AT-----TPKP-QKTPKAPKPTSTKKPKTMRPRVRKPTTTPRKMSTNPELNPTS 973
DB 660 TILKPKQPKAPKPTKPKKPTKAPKPTSTKKPKT-PKTRKPKTTPAPLKTTSATPELNTP 718
QY 974 RIAEAMLOTTTRNQTNSKLVNPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 1033
DB 719 --LEVLPPTTTPKQTPNPETAENVNPDHEDADGEGEKP-LIPGPPVLFPTAIPGTDLLA 775
QY 1034 RVNNOGIIINPMLSDETNICNGKVPDGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRIT 1093
DB 776 GRNLRGINIMPSEDETNLNCGRPVDGLTTLRNGTLVAFRGHYFWMLNPPSPSPARRIT 835
QY 1094 EVWGIPSPIDTVTRCNCEGKTFEFDKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 1153
DB 836 EVWGIPSPIDTVTRCNCEGKTFEFDKDSQYWRFTNDVDPGPKQIVKFGGLTGKIVAA 895
QY 1154 LSTAKYNWPSVYFFKRGSGSQYIYKQEPVQKCPRRPALNYPVYGMTOVRRRFE 1213
DB 896 LSTAKYKDRPESVYFFKRGSGSQYIYKQEPVQKCPRRPALNYPVYGMTOVRRRFE 955
QY 1214 AIGPSQTHTRIOVS-PAFLAYQDKGLHNEVKVSLWRGLNPNVVTSAISLPIRKPDGY 1272
DB 956 AVGFQTHTRIHVSVMRVSQDKGLHNEVKVSLWRGLNPNVVTSAISLPIRKPDGY 1015
QY 1273 DYAFSKDQYNNIDVPSTARAITTRSGQTLKSKWYNCP 1311
DB 1016 DYAFSKDQYNNIDVPSTARAITTRSGQTLKSKWYNCP 1054

```

RESULT 6

```

O77765 PRELIMINARY; PRT; 401 AA.
AC O77765;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99120896; PubMed=9920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
RA Kuetner K.E., Caterson B.;
RA "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT proteoglycan with potential growth-promoting, cytoprotective, and
RT lubricating properties in cartilage metabolism.";

```

RESULT 7

```

MUC2 HUMAN
ID MUC2 HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RA "Molecular cloning of the human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";

```

RL Biochem. Biophys. Res. Commun. 254:535-541 (1999).

[2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Articular cartilage;
 RX MEDLINE=99171663; PubMed=10073655;
 RA Schumacher B.L., Hughes C.E., Kuetner K.E., Caterson B.,
 RA Aydelotte M.B.;
 RA "Immunodetection and partial cDNA sequence of the proteoglycan,
 RT superficial zone protein, synthesized by cells lining synovial
 RT joints.";
 RL J. Orthop. Res. 17:110-120 (1999).
 DR EMBL; AF056218; AAD13404.1; -;
 DR InterPro; IPR000585; Hemopexin.
 DR Pfam; PF00045; Hemopexin; 2.
 DR SMART; SM00120; HX; 2.
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
 SQ NON_TER 1
 FT SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

Query Match 24.3%; Score 1714.7; DB 2; Length 401;
 Best Local Similarity 80.2%; Pred. No. 1.3e-28;
 Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

```

QY 914 PKDRATNSKATTPKPKQKPTKAPKPTSTKKPKTMRPRVRKPTTTPRKMSTNPELNPT 972
DB 5 PKGRATNSQVTTTPKPKQKPTKAPKPTSTKKPKT-PRVRKPKTTPPKTTTSAMPBPTPT 63
QY 973 SRTAEAMLOTTTRNQTNSKLVNPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 1032
DB 64 S-LPEAMLOTTTRNQTNSKLVNPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 122
QY 1033 PRVNOGIIINPMLSDETNICNGKVPDGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRI 1092
DB 123 VRGSGQFGINPWFSEDETNLNCGRPVDGLTTLRNGTLVAPRGHYFWMLTPTTPPPPRRI 182
QY 1093 TEWVGIPSPIDTVTRCNCEGKTFEFDKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 1152
DB 183 TEWVGIPSPIDTVTRCNCEGKTFEFDKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 242
QY 1153 ALSTAKYNWPSVYFFKRGSGSQYIYKQEPVQKCPRRPALNYPVYGMTOVRRRFE 1212
DB 243 ALSIAQKSPESVYFFKRGSGSQYIYKQEPVQKCPRRPALNYPVYGMTOVRRRFE 302
QY 1213 RAIGPSQTHTRIOVS-PAFLAYQDKGLHNEVKVSLWRGLNPNVVTSAISLPIRKPDGY 1272
DB 303 RAIGPSQVHTIRIHVTPRVYQDKGLHNEVKVSLWRGLNPNVVTSAISLPIRKPDGY 362
QY 1273 DYAFSKDQYNNIDVPSTARAITTRSGQTLKSKWYNCP 1311
DB 363 DYAFSKDQYNNIDVPSTARAITTRSGQTLKSKWYNCP 401

```


[illegible]


```
Db 1031 TLKTEBPTTRKTSFTTKTREPPTTRKRVTTERTTREPPTSKTTHHTTAEPATKTTHEP 1090
Qy 659 TTPKGTGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKPEPA--- 697
Db 1091 TTQKSTTLRITBEPITRKSSTAKTTRPTTKRETTERTTKEPTTRKTTTHKTTBEPPTKK 1150
Qy 698 -----PTTPK-----KPAPKELAPTTPKGTSTTSKDPAPTTPKETAPTTPKEPA----- 742
Db 1151 TTTEPTTKSTTLKPTEE---PTTRK---TSTTKTTRPTTKKTTERTITKEPTTRKTTTH 1205
Qy 743 -----PTTPK-----KP-----APTPEPPPTTSVSPPTTK 771
Db 1206 KTBEPPTNKKTKTTEPTTKKSTLAPTEPTTRKTSITTKTREPPTTKVTERTR 1265
Qy 772 EP-----TTHKSPDE-----STPELSAEPPTKALENSPKEGVPPTTKPAATKPEMTT 821
Db 1266 EPTTRKTTTHKTEBPTTKKTTTKTTEPTTK-STTLKPTBEPPTTRKTSITTKTREP 1324
Qy 822 AKDKTERDLR-----TTPETTTAEP----- 842
Db 1325 TKRVITERTTRBPTTRKTTTHKTEBPTTKKTTTKTTEPTTKSTTLKPTBEPPTTRKT 1384
Qy 843 ----KTKETAT---TTEKTESKITATTQVTSITTTQDTPPKITTLKT-----TT 887
Db 1385 STTKTTRPTTRKRVITERTTREPPTTRKT--THKTTBEPPTTKKTTTKTTEPTTKSIT 1442
Qy 888 LAP-----KVTTK-----KITTTTEIMNKP-----BETAKPKORATNSKAT 925
Db 1443 LKTEBPTTRKTSITTKTREPPTTRKRVITERTTREPPTTRKTTTHKTEBPTTKKTTTKIT 1502
Qy 926 PKPQ-----KP-----TKAPKKPT----- 939
Db 1503 HEPTTKKSTTLKPTBEPPTTRKTSITTKATREPPTTRKRVITERTTREPPTTRKTTAKHTTEPT 1562
Qy 940 -----STKKPKTMVRVKPTTPTRPKMTSTPELNPTSRIAEAMLQIT--- 983
Db 1563 TKTKTKTKTTEPTTKKSTTL-----KTEBPTTRKTSITTKTTEPTTRSTSVKITADQ 1618
Qy 984 -----TRPNQTPNSKLAVNPKSSEDAGGAETP-----HMLLRPHV 1020
Db 1619 TTKRTTAEMSTTNOEPTSVETTTNSSQNTTSTTTEBQVHHHHHHIHYH---KPAD 1675
Qy 1021 FMEVTPDMYLRVFNVOGIIINPLMSDEINI CNKPKVDGLTTLNGLTVAFRGHYFWML 1080
Db 1676 LGPSILP-----LPDLP----- 1687
Qy 1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCEGKTFEFDKDSQYWRFTNDKAGYKPIF 1140
Db 1688 -PLPLPLP-----WP-PLPLPEI-----PLPL- 1707
Qy 1141 KGFGLTGTOIVAALSTAKYNWPESVYVFKRGSGIOQYIVYKQEPVQKCGRRPALNVPV 1200
Db 1708 -----PPLPTA-----LPPLPLPLPP----- 1725
Qy 1201 GEMTOVRRRFRFAIGFSQHTIRIQYSPARLAYQDKVLHNVKYSILWGLNPNVTS 1260
Db 1726 -----LPEVNLT 1733
Qy 1261 ISLPNIRKPD 1270
Db 1734 ISLPEISLPN 1743
RESULT 11
ID AAS64673 PRELIMINARY; PRT; 1761 AA.
AC AAS64673;
DT 01-APR-2004 (TrEMBLrel. 27, Created)
DT 01-APR-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE CG33300-PA (Fragment).
GN CG33300.
```

```
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer J.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
RA Arkil J.F., Aqbayani A., An H.J., Andrews-Pfannkoch C., Besley E.M.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beaskey S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Brottier P.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Talali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris D.M., Nelson D.L.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Patterson K., Saunders R.D., Scheeler F., Shen H.,
RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,
RA Shue B.C., Siden-Kiamos I., Stappleton M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Stappleton M., Turner E., Wang A.H., Wang X.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.E., Wan K., Stappleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RL melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stappleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RL systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[4]
RP SEQUENCE FROM N.A.
```


QY	1141	KGFGGLTGCQIIVAALSTAKYKNWPSVYFFKRGGSIQQVIYKQEPVQKCPGRRPALNYPVY	1200
Db	1708	-----PPLFTA-----	1725
QY	1201	GEMTOVRRRFRERAIQPSQTHIRIQYSPARLAYODKGVHLNEVKVSIILWRGLPNVVTSA	1260
Db	1726	-----LPEVNUTA	1733
QY	1261	ISLNPIRKPD	1270
Db	1734	ISLPEISUPN	1743
RESULT 12			
O7PMD5 PRELIMINARY; PRT; 3150 AA.			
AC	Q7PMD5	PRELIMINARY; PRT; 3150 AA.	
DT	01-MAR-2004	(Tremblrel. 26, Created)	
DT	01-MAR-2004	(Tremblrel. 26, Last sequence update)	
DT	01-MAR-2004	(Tremblrel. 26, Last annotation update)	
DE	ENSANGP00000004655	(Fragment).	
GN	Name-ENSANG00000003651.		
OS	Anopheles gambiae str. PEST.		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.		
OX	NCBI_TaxID=180454;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=PEST;		
RC	Anopheles Genome Sequencing Consortium;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
CC	!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
CC	EMBL; AAAB01008980; EAA13969.2; -.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0004872; F:receptor activity; IEA.		
DR	GO; GO:0008151; P:cell growth and/or maintenance; IEA.		
DR	InterPro; IPR006770; OGF_recept.		
DR	Pfam; PF04680; OGF_r_III; 80.		
FT	NON_TER	1 3150 3150	
FT	NON_TER	3150 3150	
SQ	SEQUENCE	3150 AA; 322879 MW; 3C7B3D441CE8C939 CRC64;	
Query Match 18.0%; Score 1271.2; DB 2; Length 3150;			
Best Local Similarity 19.0%; Pred. No. 1.5e-17;			
Matches 447; Conservative 128; Mismatches 455; Indels 1318; Gaps 93;			
QY	115	TKKKTPKPPVDEAGSLDNGDKVTTPTDSTTQHN-----	151
Db	227	TPRPTDSTMSA-----STPESTPGTTRTTPRPTSTESTDITMSSAST	275
QY	152	KVSTSPKITAKPINP-----RPSLPNNSDTSKE	180
Db	276	PESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMS	335
QY	181	TSLVNKEITVETKETTNNKQNSTD-----GKEKTSKAK--ETQSIKLT	223
Db	336	SASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDIT	394
QY	224	SAKDLAP-----TSKVLAKTPKAEI-----TTKGPAITTPKEPTP	259
Db	395	MSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTES	454
QY	260	-----TTKPEAS-----	273
Db	455	TDTLSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTE	514
QY	274	-----TPTTIKSAPTTP-----	300
Db	515	STDTMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPR	574

QY	301	-----KEPAPTTTKEPAPTTKPEPAPT-----TTKEPA--PTTTKSA	335
Db	575	TSSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTP	634
QY	336	PTTPKEPAPTTKPEPAPT-----TPKEPAP-----TTPKETP	368
Db	635	PTPTSTGKTTPRPTSTESTDITMSSASTP--EPSTTPGTRTTPRPTSTESTDITMSSAST	693
QY	369	-----TTKPEPAPT-----TKEPAP-----TTPKPEPAPT-----	392
Db	694	PEPSTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMS	753
QY	393	APKKDAPTTKPEPAPTTKPEPAPT-----TTKEPSP-----TTKPEPA	430
Db	754	SASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITM	813
QY	431	-----PTTTKSAPT-----TTKEPA--PTTTKSAPT-----	456
Db	814	SSASTSEPTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTEST	873
QY	457	-----KEPSTTTTKEPAPTTKPEPAP-----TTPKKPAP-----	485
Db	874	DITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTES	933
QY	486	-----TTKPEPAP-----TTKPEPA	500
Db	934	TDTMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPT	993
QY	501	-----PTTK-----	509
Db	994	STESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTP	1053
QY	510	TAP-----KEPAPTTKPEPAPTTPKKTTP-----TPE-----	537
Db	1054	TRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTT	1113
QY	538	KLAPT-----TPE-----	552
Db	1114	RTTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTP	1173
QY	553	ELAPTTPEPTP-----TPE-----	572
Db	1174	PGTTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPE	1233
QY	573	PTTPKAAAPNTKPEPAP-----TTKPEP-----	595
Db	1234	PSITPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTP	1293
QY	596	-----APTTP-----KEPAPTTKPEPAPTTPKGTATTLKE-----	626
Db	1294	EPSTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITM	1351
QY	627	-----PAP-----TTPKKPAPKELAPT-----TKEPTST-----	651
Db	1352	SSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTEST	1411
QY	652	-----TSDKPAPTTKGTATTLKPEP-----APTTP-----	677
Db	1412	DITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTES	1471
QY	678	-----KEPAPTTKGTATTLKPEP-----	702
Db	1472	TESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTP	1531
QY	703	KPAPKELA-----	710
Db	1532	REPTEITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTRTTP	1591
QY	711	-----PTTKGPTSTTSKPA-----PTPKETA	734
Db	1592	TRPTSTESTDITMSSASTPESTTPGTRTTPRPTSTESTDITMSSASTPESTTPGTT	1651
QY	735	PTTPKEPAP-----TTPKKPAPT-----TPE--TTP	758


```

Db 1652 RTTETPTDSTMTSSASTPESTPGTTRTTPRTPTDSTMTSSMSSESTPESTTPE 1711
Qy 759 PTT-----SEVSTTTTKEPTTIHKSP-----DE 782
Db 1712 PSTTPGTTRTPTSTESTDSTMTSSASTPESTTPTGTTRTPTSTESTDSTMTSSA 1771
Qy 783 STPELSAEP-----TPKALENSKEP-----GYPTT 808
Db 1772 STPESTPGTTRTPTSTESTDSTMTSSASTPESTTPTGTTRTPTSTESTDSTMT 1831
Qy 809 KTPAATKPEMTT-----TAKDKTTERDLRTP-----835
Db 1832 MSSASTPESTTPTGTTRTPTSTESTDSTMTSSASTPESTTPTGTTRTPTSTES 1891
Qy 836 -----ETTTAAKMKIKETATT-----TEKTESKIT 861
Db 1892 TDTMTSSASTPESTTPTGTTRTPTSTESTDSTMTSSASTPDSTTPTGTTRTPTPT 1951
Qy 862 ATTQVISTTQD-----TTPFKIT-----TLKTTTLAP 890
Db 1952 PTETMTSSASTPESTTPTGTTRTPTSTESTDSTMTSSASTPESTTPTGTTRTPT 2011
Qy 891 KVTTKKTIITTEIMKPBETAK-----KDRATNSKATTPKQ-----929
Db 2012 --TSTERTDSTMSAYTPESTTPTGTTRTPTSTESTDSTMTSSASTPESTTPTGTR 2069
Qy 930 ---KPTK-----APKPTSTK-----KKTMPRVR 951
Db 2070 TPTRTSTSTMTSSASTPESTTPTGTTRTPTSTESTDSTMTSSASTPESTTPTGT 2129
Qy 952 K-----PKTTPRKMTSTMPELNPT-SRIAEMLQTTTRP 986
Db 2130 RTTPTPTSTSTMTSSMSASTPESTTPTGTTRTPTSTESTDSTMTSSMSASTTP 2187
Qy 987 NOTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPML 1046
Db 2188 GTTTRTPTSTESTDSTMTSSASTPE-----PSTTPG-----2220
Qy 1047 SDETINCKGKPDGLTLNGLTIVAFRGHYFWMLSPFSPSPA-----RRITEVWGI 1101
Db 2221 TTRTPTPTPTD--STMGSS-----MSSASTPESTTPTGTTRTPT--RPTP 2264
Qy 1102 IDTVFRNCEGKTFEFDKSDQYWRFTNDIKDAGYKPIKFGGLT-----1147
Db 2265 TDTT-----MSSMSASTPEP--STTPGTTRTPTPTPTPTDSTM 2302
Qy 1148 GOIVAAALSTAKYKNWPESVYFFKRGGSIQQ-----YIYKQB- 1183
Db 2303 SSSMSASTPESSITPE-----LQQDQHLPIQCRLVCROPLLRSHRHLVQPEL 2353
Qy 1184 -----FVQKCPGRRPALNYPVYGMTQVRRRFRERALGPSQTHIRIQYSPABLAY 1236
Db 2354 LQQDQHLPIQC--RQPL-----RSHLRHLVQP-----ELQD 2386
Qy 1237 KGVLHNEVKVILMGLPNVVT--SAISLENIRKPDGYDYAFKSDQYINIDVRSARA 1294
Db 2387 -----QHLFSTDTMTSSASTPE-----PSTTPG- 2409
Qy 1295 ITRRSQGT 1302
Db 2410 -TTRTPT 2416

```

RESULT 13

Q6SSE6

ID Q6SSE6

AC Q6SSE6

DT 05-JUL-2004

DT 05-JUL-2004

DE 05-JUL-2004

DE plus agglutinin.

GN Name=SAG1;

PRELIMINARY; PRT; 3409 AA.

(TrEMBLrel. 27, Created)

(TrEMBLrel. 27, Last sequence update)

(TrEMBLrel. 27, Last annotation update)

OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadales; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
 RA Goodenough U.W.;
 RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY450930; AAS07044.1;
 DR InterPro; IPR002951; Atrophin.
 DR InterPro; IPR003882; Pistil_extensin.
 DR PRINTS; PRO1222; ATROPHIN.
 DR PRINTS; PRO1218; PSTEXTENSIN.
 SQ SEQUENCE 3409 AA; 336045 MW; 7FB87633EDD6631F CRC64;

Query Match 17.28; Score 1211.6; DB 2; Length 3409;
 Best Local Similarity 13.74; Pred. No. 3.2e-16;
 Matches 384; Conservative 167; Mismatches 447; Indels 1814; Gaps 78;
 Qy 1 MAWKTLPI-----YLLLLLSV-----16
 Db 50 LAWTIPVSDALQFPVNVSVTSSAATDAPSAIAYIGALLDLVNWVQPCINDRRSYD 109
 Qy 17 -----FVIQYSS-----ODLSSACGRGEGYSRD 41
 Db 170 DFGIYQPVLDGDFMMWGFVTNTVDSNLKWDIPAWTAQGAWLGGDF-----RD 217
 Qy 42 ATCN-----CDXNCQHYMECCDFKRVCTAELSKGRCFESFERGECDDACKKYD-- 94
 Db 218 AVNFGTHYCSW--PFVEC-----SSCENYDIA 243
 Qy 95 -----94
 Db 244 DYPDPKIANGIVPAVITALDFRNASLYIYLDGFMFAGSLDNWYMLNAYNFTGGPLP 303
 Qy 95 -----KCCPD-----YESFCAE-----106
 Db 304 ANLPTLLPSLOHLAHDHCRATEDVRGTASLOYGWQQYPTGQPYE--YCSGDGVTGSDGTE 362
 Qy 107 -----106
 Db 363 YVISGMIPDENGDAVAAASDSSLFWANLRVRLSNQALYGIPEGLRSASSISSWRILQ 422
 Qy 107 -----VKDNKKNRK-----116
 Db 423 GNTELCGLPEFAAPINSLLYLGLTGTWVHDPADHTQGECLAPPPPPSPRPP 482
 Qy 117 -----116
 Db 483 RPPPLPPSPPPPLPPSPVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 542
 Qy 117 -----KKTPKPPVVDVDEAGSLDNGDFKVTTPDT 145
 Db 543 GGICDPSMPPPSPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 589
 Qy 146 ---STQHNVKYSTGKITAKPINRPSLPNSDTSKETSILVNKETTVEKTTTNNKQ 202
 Db 590 PPASSVPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPP 620
 Qy 203 TSTDGKEKTSKETSQIEKTSKADLAPTSKVLAKPTPKAETTKGPAITTPKEPTTP 262
 Db 621 -----PPSPAPPSPAPPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 647
 Qy 263 KEPAST-----TPKPTPTTKSAPTTPKEPATTTKSAPTTPKEPATTTKEPATTT 315
 Db 648 QPSPVPPSPKPPSPAPPSPVPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 707
 Qy 316 PK-----EPAPTTTKBPATTTKSAPTTPKEPATTTTPKKPAPT-----TPKE 357


```

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadales; Chlamydomonas.
OX NCBI_taxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RT "plus and minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AA507044.1; -
SQ SEQUENCE 3409 AA; 336045 MW; 7FB87633EDD6631F CRC64;

Query Match 17.2%; Score 1211.6; DB 2; Length 3409;
Best Local Similarity 13.7%; Pred. No. 3.2e-16;
Matches 384; Conservative 167; Mismatches 447; Indels 1814; Gaps 78;

QY 1 MAWKTLPI-----VLLLLLSV----- 16
Db 50 LAMVTIPVSALQFPVNVSVTSSAATDAPSAIAYIGALLDLVWVMPQNCINDTRSYD 109
QY 17 -----FVIQVSS-----QDLSSCAGRCGEGYSRD 41
Db 170 DFGIYQPVLDGFMWVGFTVNTVDSNLKWI DPATQAQGLGQDF-----RD 217
QY 42 ATCN-----CDYNQHYMECCDFKRVCTAELSCKRCFESFERGECDCDAOCKYD-- 94
Db 218 AVWNFGTHYCSW---PFVEC-----SSCENYDIA 243
QY 95 -----KCCPD-----YESFCAE----- 106
Db 304 ANLPTLLPSLOHLAHDHCRATPDVRGTASLQYGMQYPTGTQPYE-YCSGDVGTGSDGTE 362
QY 107 -----KCCPD-----YESFCAE----- 106
Db 363 YVISGMIPDEWGDVAAASDSSLFWANLRTVRLSNQALYGIPEGLRSASSISSWRLQ 422
QY 107 -----VKDNKKNRK----- 116
Db 423 GNTELCGLPEFAAPINSLLYGLTGLGTHTWHPADHTQGECLAPPPPPSPRPP 482
QY 117 ----- 116
Db 483 RPPPLPSPPPPLLPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 542
QY 117 -----KKPTPKPPVDEAGSLDNGDFKVTTPDT 145
Db 543 GGICDPSPMPPSPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 589
QY 146 ---STQINKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKEITTTNKQ 202
Db 590 PPASSVPPSPPPSPKPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 620
QY 203 TSDGKEKTSKAKTSQIEKTSKADLAPTSKVLAKPTPKAETTTKGALATTPKPTPTTP 262
Db 621 -----PPSPAPSPAPSP-----PAPSPQPPSPVPP 647
QY 263 KEPAST-----TPKEPTPTTIKSAPTTKKEPATTTKSAPTTKKEPATTTK 315
Db 648 QPSPVPPSPKPPSPAPSPVPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 707
QY 316 PK---EPAPTTKEPATTTKSAPTTKKEPATTTKPKKAPT-----TPKE 357
Db 708 PSPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 767
QY 358 PAPTTPKEPTPTTKKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPATTT 413

```

SEQUENCE FROM N.A.
Jones A.R., Hughes C.E., Flannery C.R., Catterson B.;
Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL; AY536119; AAS83977.1; -;
InterPro; IPR000585; Hemopexin.
Pfam; PF00045; Hemopexin; 2.
SMART; SM00120; HX; 2.
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
NON_TER 1 251
NON_TER 251 251
SEQUENCE 251 AA; 28532 MW; 2011B2C1719AA68B CRC64;
Query Match 16.7%; Score 1183; DB 2; Length 251;
Best Local Similarity 87.2%; Pred. No. 8e-18;
Matches 218; Conservative 10; Mismatches 22; Indels 0; Gaps 0;
QY 1055 KPVVDGLTLRNGTLVAFRGHYFVWMLSPFSPSPARRITEVWGIPSPIDTVEFTRCNCEGK 1114
Db 1 GRPVDGLTLRNGTLVAFRGHYFVWMLTPTPTPPPPRRITEVWGIPSPIDTVEFTRCNCEGK 60
QY 1115 TPEFKDSQVWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFPRKGG 1174
Db 61 TPEFKGSQVWRFTNDIKDAGYKPIFKGFGGLTGKIVAAALSTAKYKSPESVYFPRKGG 120
QY 1175 IQQYIYKQEPVQKCPRRPALNYPVYGVEMTQVRRRFRERAIQPSQTHIRIOYSPARLAY 1234
Db 121 VQQYTYKQEPVQKCTGRRPAINYSVYGVETAQVRRRFRERAIQPSQVHTIRIHYTPVRVY 180
QY 1235 QDKGVLHNEVKVSIILWRGLPNVTSAISLPNIRKPDGYDYAFSKDQYNNIDVPSRTARA 1294
Db 181 QDKGVLHNEVKVSTLWAGLPNVVTSALSLPNIRKPDGYDYAFSKDQYNNIDVPSRTARA 240
QY 1295 IITRSGQTLS 1304
Db 241 IITRSGQTLS 250
Search completed: October 13, 2004, 12:20:38
Job time : 182.144 secs

1831 VKGNLGAFAEAMPTNEKEVTAAEQARQVGGDLWPIGNLGGSSNTTASGSSSSSSSSN 1890
QY 925 TPKEOKP-----TKAPKKPTSTK 942
Db 1891 SP-PRRPPPPAAGSTGLFSNADASPPPLAVATPALPAPLSTIAAATAAP-----1942
QY 943 KPKTMPVRKPKTKTTPPRKMTST-MPELN-----PTSRIAEAMLOT 982
Db 1943 -----PRLPSP---PPPAVGSSTGLVPRRHLMQMLQPPFAAUAAPPPPPASSALVLPQ 1994
QY 983 TTRPQNTNSKIVE-----VNPKS 1001
Db 1995 SPPEPPPSQLIQOASATYVSDMDLYTLVVAAMLITAVAGLLIAAVALYRLVSE- 2053
QY 1002 EDAGAEGETPHMLLRPFVMEVETDMDYLRVFNQOIIINPMLSDETNICNGKPVQGL 1061
Db 2054 -----PPFLA-----PPRL-----2064
QY 1062 TTRNGTLVAFRGH-----YFWMLSPFSPSPA 1089
Db 2065 TTIAGLLIIVALTFFSCMALGPAADWHGSRTAAYCVLTIIVVYAAFLWLA-----2116
QY 1090 RRIEVTWGP-----SPIDTVFTRCN-----1110
Db 2117 --LARAAMVPQTLVEPMTTSSVAPRPSAFERADRTSPRSTVSIATAADIGGTATGG 2174
QY 1111 -----CEKTEFFKDSQV---WR-FTNDIKDAGYKPIFKGFG 1144
Db 2175 GGAASACADVGAMAAPATVAGADVPSEDDRYARGPHWKQF-----DGLPPTVASGVG 2229
QY 1145 GLTGQIVAAALSTAKYKNWPESVYFPRKGGSIQQYIYKQEPVQKCPRRPALNYP-----1198
Db 2230 GSGGGVVPL-----PPLVALPGSGRHLPLPPLPAA 2260
QY 1199 -----VYGVEMTQVRRRFRERAIQPSQTHIRIOYSPARLAYQDKG 1238
Db 2261 GAATAAGGGAATAPGPRGAAGB-----DTH-----QFCP-----2291
QY 1239 VLHNEVKVSIILWR-----1251
Db 2292 --H-----WRKFSVPDGAATAGVVDAAAGAGGGGGRSGSTGVRFANGTRT 2341
QY 1252 -----GLP-----NVVT 1258
Db 2342 PSDDGAKGRSSSHGGGNGNMSGASFCASAGGGGAAAAASRLGSRSGSPGCDYNAMS 2401
QY 1259 SAISL-----1263
Db 2402 SGAMRGGGAGASRFAGSAGSPGAGGAGGAGANAFAGTAGRAGATSPAHDGSSRP 2461
QY 1264 -----PNIRKP-----DGYDYAFSKDQYNNIDV 1287
Db 2462 GSGSNEGSRQMSYGSNNAMTAGAVGVGGVRRPAGSLRNASDG-----DA 2507
QY 1288 PS-----RTARAI-----TTRSG 1300
Db 2508 PPLPAPGTVPLPFANLLTSGADVRSRLNAEAVAAAAAGGGGGGTSSG 2559
RESULT 15
Q6QJF3 PRELIMINARY; PRT; 251 AA.
AC Q6QJF3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Proteoglycan 4 (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 26.7037 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKTLPIVILLLSVFIQ.....ARAITRSGQTLSKVWYNCP 1311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1382.7	19.6	3020	2	A43932		mucln 2 precursor,
2	1154.2	16.3	1664	2	T18262		S-layer protein -
3	1038	14.7	7962	2	I38346		elastic titin - hu
4	1026.6	14.5	3570	2	T45025		mucln MUC5B, trach
5	1021.4	14.5	2187	2	T30826		cyst germination s
6	1011.3	14.3	1489	2	T31108		nascent polypeptid
7	1010.5	14.3	1274	2	T16251		hypothetical prote
8	956.2	13.5	1367	1	S48478		glucan 1,4-alpha-g
9	951.7	13.5	3507	2	T29757		hypothetical prote
10	945	13.4	6642	2	T34513		protein UNC-89 - C
11	900.4	12.7	1188	2	S49915		extensin-like prot
12	895.9	12.7	4135	2	T42629		proline-rich pepti
13	885.5	12.5	5762	2	A18189		tenascin-X - bovin
14	870.7	12.3	1229	2	T25697		hypothetical prote
15	867.5	12.3	1344	1	A35175		mucln 1 precursor,
16	841.9	11.9	5262	2	T03454		ALR protein - huma
17	837.8	11.9	2897	2	B48666		cell proliferation
18	837.8	11.9	3256	2	A48666		cell proliferation
19	824.6	11.7	3942	2	T42730		Bassoon protein -
20	823.4	11.7	1151	2	T18535		high molecular mas
21	821.3	11.6	2232	2	T34434		hypothetical prote
22	814.8	11.5	1832	2	T31113		mucln-like glycopr
23	809.4	11.5	3938	2	T42761		Bassoon protein -
24	808.9	11.5	2142	2	B35038		MHC class III hist
25	805	11.4	4957	2	T03455		apoptogen(a) (EC
26	800.6	11.3	4548	1	S00657		MHC class III hist
27	798.7	11.3	1872	2	S36152		MHC class III hist
28	793.3	11.2	1870	2	S37671		probable tenascin
29	787.2	11.1	4006	2	T09070		

ALIGNMENTS

RESULT 1

A43932
mucln 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU>
A;Cross-references: UNIPROT:002817; UNIPROT:Q14884; GB:I21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence incontinent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33332; MUID:89197956; PMID:2703501
A;Accession: B33332
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

30	783.2	11.1	2225	2	T26063	hypothetical prote
31	778.9	11.0	3381	2	T42389	versican precursor
32	772	10.9	1630	2	A53577	ascites sialoglyco
33	765.7	10.8	1611	2	T38236	hypothetical prote
34	758.8	10.7	3566	1	A40701	tenascin-X precurs
35	756.1	10.7	5105	2	T32650	hypothetical prote
36	752.6	10.7	1777	2	T34369	hypothetical prote
37	751.2	10.6	761	2	C84672	hypothetical prote
38	749.2	10.6	839	2	F75518	hypothetical prote
39	748.2	10.6	2774	2	A43359	microtubule-associ
40	747.4	10.6	3149	1	Q0888	BRF1 protein - hu
41	746.3	10.6	924	2	S27923	gene Lf3 protein -
42	741.7	10.5	990	2	I51618	nucleolar phosphop
43	735.3	10.4	5170	2	T15348	hypothetical prote
44	735.1	10.4	971	2	T19431	hypothetical prote
45	725.8	10.3	4667	2	T20774	hypothetical prote

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481; PMID:1985113
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R:Yu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
A:Reference number: P00328; MUID:92198477; PMID:1550588
A:Accession: P00328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M86523
A:Experimental source: small intestine
A:Accession: P00329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: lip15.5-lip15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
C:Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 19.6%; Score 1382.7; DB 2; Length 3020;
Best Local Similarity 18.5%; Pred. No. 2.9e-21;
Matches 482; Conservative 126; Mismatches 567; Indels 1433; Gaps 89;

QY	6	LPYILLLSVFVIQVSSQDLSACAGCG	-----EG-----	YSR	40
DB	502	LQVLAPVQLFVTLDAQSQ	-QVQGLCGNFNLEGGDFKTASGLVEATGAGFANTWKA		559
QY	41	DATCN	-----CDY-----		48
DB	560	QSTCHDKLDWDFCSINIESANYAHCWSLLKKTETPGRCHSAVDPAEYKRCYDTC			619
QY	49	NCOHYMEC	---CPDFKRVCTA---	ELSK	71
DB	620	NCONNEDCLCAALSSVARACTAGWLMGWREHVCKNDVGCNSQVFLYNLTCCQTCR			679
QY	72	-----GRCPESF-----	ERGR-----ECCD-----	DAQCKYDK	95
DB	680	SLSEADSHCLEGAPVDCGCPDHTFLDEKGRCVPLAKCSYHRLGLEAGDVVVRQER			739
QY	96	C	-----		96
DB	740	CVCRDGRLHCRQIRLIQGSCTAPKIHMDCSNLTALATSKPRALSCOTLAAGYHTECVSG			799
QY	97	---CPD---	---YESFC---	AEVK---	108
DB	800	CVCEDGLMDGGCGVCEKPCVHNNDLYSSGAKIKVDNCTCKEGRWVCTQAVCHGT			859
QY	109	-----DNK-----			111
DB	860	CSYSGSHYITFDGKYVDFDGHCSYAVQVDCGQNSLGSFSLITENVPQGTGVCSCA			919
QY	112	-----KNRTKKPTPKPV-----			133
DB	920	IKIFMGRTELKEDKRVVIQRDEGHHVAYTRVGVLYVVESTGLIIVWKKRTTFIK			979
QY	134	-----DNGDFKV-----			149
DB	980	LAPSYKGTVCGLCGNFDRSNNDFTRDHMVSSSELDFGNSWKEAPTCPDVSTNPEPCL			1039
QY	150	-----HNKVTSP-----			157
DB	1040	NPHRSWAECQCSILKSSVFSICHSKVDPKPFYEAHVHDSGSCDGTGSGCEFCSAVASYA			1099
QY	158	-----KITAKPIN-----			166

Db	1100	QECTKEGACVWRTPDLCPICDYVNPHECEWHYPCGNRSFETCRTINGIHSNISVSY	1159
QY	167	-----PRPSLP-----	172
Db	1160	LEGCVPRCPKDRPIVEDLKKCVTADKCGCYVEDTHYPFGASVPTTECKSCVCVNSSQV	1219
QY	173	-----PNSDTSKE-----	194
Db	1220	VCRPEGKILNQDQGAFCYWEICGPNGTVEKHFNCISITTPSPSLTITFTITLTPPTS	1279
QY	195	ETTTINKOTSDGKEKTTSAK	223
Db	1280	FTTTTTTPTSSVLSLSTPKLCLLWSWINEHPSSGSDGDRFFDGVCGAPEDIECR	1339
QY	224	SAKD	231
Db	1340	SVKDPHLSLEQHGQKQVQCDVSVGFICKNEDQPGNPGFLCYDYKIRVNCWPMKDCIITP	1399
QY	232	SKVLAKPTPKABTTTTKGPALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAP	290
Db	1400	SPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPLTTPSPPTT	1459
QY	291	TTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP	341
Db	1460	TTTTTPPTTTPSP-PTTTPSPPTTTPSPPT-TTTTTPPTTTPSPPTTTPSPPTTTP	1517
QY	342	PAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT	401
Db	1518	PPTTTPSPPTTTTTPPTTTPSPPTTTTTPSTTTTTPPTTTPPTTTPPTTTPPTTTP	1577
QY	402	BKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT	461
Db	1578	PPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSP-PTTT-PTTP-PTTP-PTTP	1629
QY	462	TTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP	520
Db	1630	TSITTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTP	1689
QY	521	KETAPTTPKLLPTTPEKLAAPTTPKEPAPTTPBELAPTTPPEPTTPPEPAPTTPKAAA	580
Db	1690	-----SPPTT-----	1735
QY	581	PNTPKAPA-----PTTPKEPAPTTPKEPAPTTP-----KETAPTTP-----	616
Db	1736	--TPSPPTTMTTLPTTSSPLTTLPLPSIPTTSPPTTTPPTTTPCVPCLCNWTCGLD	1793
QY	617	KG-----	618
Db	1794	SGKPNFHKPGGDTLIGDVGCGWAANISCRATWYDPVPIGOLGQTVVCDVSVGLICNE	1853
QY	619	-----TAPTTLKEPAPTTPKPAKELAPTTP-----TKBPTS	650
Db	1854	DQKPGVIMAFCLNYEINVQCCVQPTTM-----TTTTTENPTPTTITTTTPTPTP	1910
QY	651	TTSDKP-----APT-TPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKKA	705
Db	1911	TSQSPNGLQAPPTPISITTTTPTPTGTQPT-TTPTTTTTPPTTPTPTGTQPT	1969
QY	706	PKELAPTTPKGTPT-TSDKAPTTPKEPAPTTPKEPAPTTPKAP-----APTPTPT	757
Db	1970	TVLITTTTMTPTPTPTSTKSTVPTITTTTIVTATPTGTQPTTPTTPTTPTTPTTPTTPTT	2029
QY	758	PPTTSEVSTPTTK-----EPTTIHKSPTDSTPSELSAETPKALENSPKPGPVTTKTPA	812
Db	2030	TPTTTGTGPPHTSTAPIALTTNSPPESSTPTSTSTSPSTSTSTSTSTSTSTSTSTSTST	2089
QY	813	ATKEMTTTAKDKTTERDLRTPETTTAAKMTKETAPTTEKTESKITAQTQVSTTT	872
Db	2090	STAPSTPTPTTSSGGHTLSPPTTTPPTTTPGTPTGTGTT--GSSSAPTPTSTVQTITSA	2147
QY	873	QDTPFKITT--LKTITLAPKVTITTKTITITTEIMNKPEE-----	910


```

QY 234 -----VLAKPTPKAEITTKGPAITTPKE----- 256
Db 5537 ELTESQYNFRNDGRCHYMIHNVTDPDDEGVYSVIARLEPGEARSTAEALVLTITKEIKLE 5596
QY 257 -----PPT-----TPKE----- 264
Db 5597 LKPPDIPDSRPIPTMPIRAVPPBEIPVWAPPVPLLPTEPEKKPPPKRIEVTKKAVK 5656
QY 265 -----PASTTPKE-----PPTTI----- 278
Db 5657 DAKKVAKPBMPTREBIVKKPPPTTLIPAKAPEIIDVSSKABEVKIMTITRKVEQKE 5716
QY 279 ----- 278
Db 5717 KEAVYEKKQAVHKEKRVFIESPEPYDELEVETPEFQPYEEDDEDYEBEIKVEAKKE 5776
QY 279 ----- 278
Db 5777 VHEEWEEDFEEGOEYEREEGYDEGBEEWEAEOEREVIQVQKEVVEESHKRVPAKVP 5836
QY 279 KGAPTTPK----- 286
Db 5837 KKAAPPKPKVKKVIEKIEKTSRMBEEKVQVTKVPEVSKKIVPQKPSRTPVQEEVIEWK 5896
QY 287 ----- 295
Db 5897 VPAVHTKKWVISEKMFASHTEEEVSIVTPEVQKEIVTEEKIHVAVSKKVEPPP-----K 5952
QY 296 APTTPKEPA-----PTTKKEPAPTTPKEPAPTTPTKSAPTTPKEPA----- 343
Db 5953 VPPELKEKPAEEVAPVPIPKKVEPPAPKVPV-----KKPVPEKKVPVPPKKEFAAP 6007
QY 344 --PTTPKKEPA-----TTPKE--PAPTTPKE 365
Db 6008 KVPVPPKKEVPBEKIPVFAKKEAPPAKVPVQKGWVTEEKITIVTORESPPPAVPEI 6067
QY 366 PTEPTTPKE--PAPTTPKEPAPTTPKEPAPTTPKAP----- 399
Db 6068 PKKKVPEERKVPVKEEYVPPPKVPA--LPKKVPPEEKVAVPVFAKAPPPRAEVSKK 6125
QY 400 ----- 399
Db 6126 TVVEEKRFVAEBKLSFAVPQVVEVTRHEVSAEBSYSBEEGVSISVYVEREBEEREA 6185
QY 400 -----TTPKEPA----- 406
Db 6186 EVTEYEVMEEPYVVEEKHLIISKRVEABAEVTERQEKIVLKPAPAKIEBPPPAKV 6245
QY 407 PTPKPE-----PAPTTPKE--PSPTTPKBP--APTTPKSAPTTP--KEPA----- 446
Db 6246 PEAPKKIVPEKKVPAPVPPKKEKVPVPPKVPPEEKVPKVPKVKMEEDLPKAVTEKH 6305
QY 447 -----TTPKSAP--TTPKPSPTTPKE-----PAPTTP----- 471
Db 6306 MOITQEEKVLVAVTKEAPPKARVPEEPKRAVPEEKVLKPKREBEPKAVTEFRKRV 6365
QY 472 -----PK-EPAP----- 477
Db 6366 KEEKUSIEAPKREPOPIKEVTIMEKEKAYLTLEEAVSQREBEEYEEYDYKEFEYE 6425
QY 478 -----TTPKKPAPT--TPKBPAPTTPKE----- 498
Db 6426 PTEYDQVEEYEREVEEHEEYITEPEKIPVKKVPVEBPVTPKAPPAKVLKAVP 6485
QY 499 -----PAPTTPKAPAPKEP----- 515
Db 6486 BEKVVPVPIPKKLKPPPKVPPEPKKVFBEKIHISITKREKQVTEPAKVPMPKRVAV 6545
QY 516 -----APTTPKETAP-----TTPK 529
Db 6546 EKVVPVKEVAPPVVPVPEKLEFEVFAVEEVVTHVEEYLVVEEBEYIHEEBEFITEE 6605

```

```

QY 530 KLPTTPTEKLAPTTPKEPAP-----TTPBELAPT----- 558
Db 6606 EVVPEVIVPVK--PEVPRXPVPEEKKVPVPPKKKEAPPAKVPVPEVPKKPEKVPVLIPKKKEP 6664
QY 559 ----TPEERTPTTPKEPAPT-----TPKAAAPNTPEKEPAPTTPKEPAPT-----TPKE 602
Db 6665 PPAKVPEVPKVPVPEEKVPVPPKKEAPPAKVPVPEVKKFVP--EKKVPVPAKKEVAPPA 6723
QY 603 PAPTTPKETAPTTPKGT--APTTLKEPAPTTPK----- 634
Db 6724 KVEVPVKLLIPEEKKPTFPVKKVEAPPKVPKKEPVPVVALPOEEVILPEEBEIVPEEB 6783
QY 635 ----- 634
Db 6784 VLPEEEVLPPEEBEVLPEEBEVLPEEBEIPPEEBEVPPEEYVPEEBEVPPEEBEVLPEVK 6843
QY 635 ----PAP-----KELAP-----TTP 645
Db 6844 PKVPVPAVPEIKKKVTEKKVVIKKEEAPPAKVPVPEVPKKVEKRIILPKKEEVLPEVT 6903
QY 646 KEP-----TSTTSKDPAPTTPKGTAPT----- 668
Db 6904 EPEEPEPIGEEIPEBPPPSIEEVEVAPRVPVVIKAVPEAPTTPVKKVEADPAKVSKK 6963
QY 669 -PKE--PAPTTPKEPAP----- 682
Db 6964 IPEEKVPVQVQKEAPPAKVPVPEKKVLPKKEAVPPAKGRTVLEEKVSVAFRQE 7023
QY 683 ----- 682
Db 7024 VVVKERLELEVVEAEVEEIPPEEBEFHVEYFEEGFHEVEEFIKLEOHRVESEHREKV 7083
QY 683 ----TTPKG-----TAPTTL--KEP--APTTPK----- 703
Db 7084 HRVIEVFAEEVEVEPEKPAKPPKGPPEISEKIIIPPKPPPKVPVPRKEPPPAKVPVKKIV 7143
QY 704 ----PAPKELAPTTTGTSTTSKDPAPTTPKETAP-----TTPKEPAPTTPK 748
Db 7144 EKVVPVPEPRVPPTKV-----PEVLPPKEVVPPEKKVPVPPAKKEAPPKVPPEA 7193
QY 749 PAPTTPKE--TPPTTSEVSTPTTKE-----PTTIHKSDESTBELSAEP-- 791
Db 7194 PKEVVPPEKKVPVPPKKEV--PPTKVPEVPAKAAVEKKVPEAIPPKPSPPPEVEEPEE 7252
QY 792 TPKALENSPKPEGVPTT-----KTPAA--TKPEMTTTAKDKTTTERDLRTPE 836
Db 7253 SPSA--PPKKEVPPVVPVPEVPEVPEKKVPAAPPKKEV-----TPV 7294
QY 837 TTTAAPKMTKETATTEKTTTTSKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTTK 896
Db 7295 KVPPEAPK-----EVVPEKKV-----PVPPPKKPEVPPPTKVPEV-----PKVAVPE 7334
QY 897 KTIITTT-----EIMNKPEETA----- 912
Db 7335 KKVPEAIPPKPSPPPEVEFEVEEVALPEEPPAEVVEPEPAAPQVTPPKNPVPEKAP 7394
QY 913 -----KPKDRATNSKATTPK-- 927
Db 7395 AVAKKPELPVVKVPEVPEVPEKKVLPVPPKPEADPAKVPVPEVPEVPEKKVAVPEK 7454
QY 928 ----POKTKAPKPTSTTKP----- 944
Db 7455 PEVPPAKVPVPEKKVPLEEKPAVVPVPERAESPPPEVVEEPEEIAPEEBEIAPEEBEKPVPA 7514
QY 945 -----KTMPRVRKKTTP----- 957
Db 7515 EEEBEPVPPAVPEBPKKIIPKVKVPVKKPEADPPKPEPEKPEKVIKPKLPRPPPPPPA 7574
QY 958 -----TPRKWTS----- 964
Db 7575 PPKEDVKEKIFQLKAIIPKKKVPENPVPEKVELTPLKVPGEKKVKLLPERKEPEKEV 7634
QY 965 -----TMEPNPTSRIAEAMLQTTTR-----PNQT 989

```

```
Db 7635 VLKSVLRKRREBEPKPKLEKVKDAVPEPPPKPVEVEVETVTKRERKIPEPTKV 7694
Qy 990 ENSK-----LVEVNPKEADAGBETPHMLLRHVFMEVTP-----1027
Db 7695 PEIKPAIPLPAPPKPE-----AEVKT-----IKPPVPEPTPIAAPVTVVVGKAE 7746
Qy 1028-----DMDYLPVYVNGIINPMLSDETNICKGKPDVGLTTLRNGTLVAFRGHYFWMLSPF 1083
Db 7747 KAPKEAAKPKGPIKGV-----PK 7765
Qy 1084 SPPSP-----ARITEVWGPIDIV-FTRCNCEGKTFPKDSQY-----WRFTNDIKD 1132
Db 7766 KTPSEIARERKLKRGSGEKKPDEAPPT-----YOLKAVPLKFVKEIKD 7810
Qy 1133 AGYKPIPKGFGGLTQCIIVA-----ALSTAKYKNWPESVYFFKRGSGIQYIYKQEPVKQCP 1189
Db 7811 ILITSEFVGSSAIECLVSPSTAIT-----WMKDSNI-----RESP-----7849
Qy 1190 GRRPALNPVYGMTQVRRRF-----EPAIGSQTHITIRIQSPA-----RLAYQDK 1237
Db 7850-----KHREIADGDKRKL-----HIIDVQLSDAGEYTCVLRGNKEK 7886
Qy 1238-----GVLHNEVKV-----SILWRGLPNVVTSAISLEN 1265
Db 7887 TSTAKLVVEELPVRFKTLEEVTVVKGQPLVLSCELNKERDVWR-----7932
Qy 1266 IRKPDGYDYAPSKQYVNIIDVPSRTA-----RAIT 1296
Db 7933-----KDGKIVVEKGRIVPGVIGLMRALT 7957

RESULT 4
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: 222899; MUID:97166151; PMID:9013550
A:Accession: T45025
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:272496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match 14.5%; Score 1026.6; DB 2; Length 3570;
Best Local Similarity 13.5%; Pred. No. 2.2e-13;
Matches 472; Conservative 123; Mismatches 531; Indels 2374; Gaps 106;

Qy 23 SSQDLSSCAGRCG-----EGYS-----RDATC-----44
Db 172 SAFGTTTCQPCQWTEWDEDPKSEQLGGDVESYDKIRAGGHLCQPKDIEQAESFP 231
Qy 45 -----NCD-----YNCQHYMECCPD-----59
Db 232 NWTLAQVGQKVCVDVHFGLVCRNWEQGVFKMVCNYIRVLCCSDHCRGRTTPPTTE 291
Qy 60 -----59
Db 292 LETATTTTQALFSTPQPTSSPGLTRAPPASTTAVPTLSEGLTSPTYTSLGTATCGPR 351
Qy 60 -----59
Db 352 QSAGSTETVGVATSLPLTRSAIPGTTGSLGTWRSPQPTLAPTTWATSPARTGTAST 411
Qy 60 -----FKRVCTAELS-----CKGR-----74
```

```
Db 412 ASKEPLATTLTSLTSELSTQAEETSPRTETTTWSPLNNTTTTSGQTTTCQPKCEWTEWFD 471
Qy 75 -----FESFER-----GREC-----DCDAQ-----89
Db 472 VDFPFGVAGDMETFEINIRAGGKWCWAPKSIIECAENYPEVSDQVQVLTCSLETGL 531
Qy 90 -CKKYDK-----CCPDYESFCAEVKDNKNRKTCK-----KPTPKP 123
Db 532 TCKNEQDQGRFNCMFCNVYVRLCCDDY-SHCPSTLATSTATSPSTPGTWTLLTPTTTA 590
Qy 124 PVWDAGGLONGDKPVT--TPDSTTTOHKNVSTSPKITTAK-----PINPSPSP 173
Db 591 TTTASTGSTATASSTQATAGPHVSTT-----ATTPTVTSSKATFSSPGTATALPALS 645
Qy 174 NSDTSKETSILT-----VNKETTVTKET-----TTTNKQT 203
Db 646 TATTTATSTFTAI PSSSLGTWRLSQTTTMAWTMATSTPSTPETVHTSTVLTITATT 705
Qy 204 STDGKEKTTSAKETOSIEKTSADLAPTSKVL-----AKP-----238
Db 706 GATGSVAITPS-----STPGTAHTTKVLTITTTTGFTATPSSSPGRARTLPVWIST 755
Qy 239 -----TPKAETTTKG--BALTTPEKPT 258
Db 756 TTTTTRGSTVTPSSIPGTHHTPTVLTITTTTVAATSSSTQTSCTGPPSLTT---T 811
Qy 259 PTT-----PKEPASTTPKEPTTIKISAP-----TTPEKAPPTTKS 295
Db 812 AITITATGSTTNPSSTPGTPPIPPVLTITATTPAATSSIVTPSSALGTHHTPPVNTAT 871
Qy 296 -----APTTP-----KEPAPTTTKEPAPT-----PK 317
Db 872 THGRSLSPSSPHTVCTAWTSAATSGILGTHITEPSTGTSHTPAATGTTCHSTPALSSPH 931
Qy 318 EPAPTTTKEPAPTTIKSAPTT-----PKEP--APT-----346
Db 932 PSSRTTESPPSGTTTTPGHTTATSRATATATPSKTRTSTLLPSQTSAPITTVTMGCEP 991
Qy 347 -----346
Db 992 QCAMSEWLDYSYPMPOGSGDPTYSNIRAAGAVCEQPLGLECAQAQGVPLRELQV 1051
Qy 347 -----PKKPAPTTTKEPAPT-----361
Db 1052 VECSLDFGLVCRNREQVKMCFNVEIRVFCNCGHCPSTPATSTATSPSTPGTWTIL 1111
Qy 362 -----TPKEPTTTPKEPAP-----TTKEPAPT-----TEKEPAPTAPKKA 398
Db 1112 TELTTTATTESTGSTATPTSLRTAPPKVLTTTATPTVTSSKATPSSPGTATALPA 1171
Qy 399 -----PTTPK-----EPAP-----TTPEKAPTTTKEPSP-----423
Db 1172 LRSTATTPATSVTPIPSSSLGTWRLSQTTTPTATMTSTATPSTDETAHTSVLTATA 1231
Qy 424 -----TTPEKAPTTTKSAPTTTKEPAPTTT-----KSAP-----TTP-----456
Db 1232 TTTGATGSVATPSSTPGTAHTTKVPTTTTGTATPSSSPGALTLPVWVSTTTTPTTRG 1291
Qy 457 -----K 457
Db 1292 STVTPSSIPGTHHTATVLTITTTTVAATGSMATPSSSTQTSCTGTPPSLTITATTATGSTT 1351
Qy 458 EPSPTTTKEPAP-----TTPEKAPPT-----TPKKEPAPTTTKEPAPTTP-----496
Db 1352 NPSSTPGTRIPVPLITTTATTPAATSTVTPSSALGTHHTPPVNTTATTHGRSLSPSS 1411
Qy 497 -----KEPAPTTTKEPAPT-----APKE 514
Db 1412 HTVKTAMTSATSGTLGTHITEPSTGTSHTPAATGTTQHSTPALSPHSSRSTTESPPS 1471
Qy 515 PAPPTP-----KETAPTTPK-----LPTTPEKLAPTT-----544
Db 1472 PGTTTTPGHTATSRRTATATPSKTRTSTLLPSSPTS-APITTVTMGCEPOCAMSEWLDY 1530
```

QY 545 -----EXP----- 547
Db 1531 SYMPGSGGDFDYSNIRAGAVCEQLGLECRAQPGVPLRELQOVVECSLDFGLV 1590
QY 548 -----APTPEELAPTTEE----- 562
Db 1591 CRNREQVKFCNFVEIRVFCNCGHCPSTATSPSTPGTWTWLLTQTTAAIT 1650
QY 563 -----PTPTPEPAP----- 585
Db 1651 ATTGSTALPSSPGTAPPPKVLTSQAATPTATSSKATSSSPRTATLPVLSTATKSTA 1710
QY 586 -----EPAPT----- 606
Db 1711 TSFTIPSSLTGTSQNRPPHPMATMSTIHPSSTPETHTSTVLTKAITTRATSSMS 1770
QY 607 TPKEAPT-----PKGTAPT-----LKEAPT----- 630
Db 1771 TPSSTPGTWTWLLTAAALPHGTPSSPGTWTWLLTPESTTATVTVPTGSTATAS 1830
QY 631 -----TPKK-----PAKELAPT----- 643
Db 1831 STRATAGLKVLTATPTTVISSRATPSSPGTATALPALRSTATTPTATSVTAIPSSS 1890
QY 644 -----T----- 644
Db 1891 LGTAWRLSQTTPTATMSTATPSTPETVHTSTVLTTATTTRGCVATPSTPGTAHT 1950
QY 645 TKEPSTT-----SDKAPT----- 660
Db 1951 TKVPTTTTGTATPSSSPGTALTPVWISITTTTTRGSTVTPSSIPGTHTAIVLTT 2010
QY 661 -----PKGTAPT----- 676
Db 2011 TTVATGSMATPSSSTQSGTPPSLTATTATTATATGSTTNPSSTPGTTPPVLLTATT 2070
QY 677 -----PKEP-----APT----- 683
Db 2071 PAATSSVTPSALGTHTPVPNTTATTHGRSLPPSPSHVPTAWTSATSGILGTHIT 2130
QY 684 -----TPKGAPT-----LKEPAP-----TPPKPAKELAPTITTKGPTSTSDKPA 726
Db 2131 EPSGTSTHTPAATTQSTPALSSPHSPSTTESPPSP-----GTTTGHTRGSTRTA 2186
QY 727 PTTKETAPT-----PKEP----- 741
Db 2187 TATPSKRTSTLLPSSPTSAPITTVVTCBPQCAWSEWLDYSYMPGPGSGGDFDYSNI 2246
QY 742 ----- 741
Db 2247 RAAGGAVCEQLGLECRAQPGVPLRELQOVVECSLDFGLVCRNREQVKFCNFVEI 2306
QY 742 -----APTPKKAPTTPPTPT----- 761
Db 2307 RVFCNCGHCPSTATSPSTPGTWTWLLTKLTATTTESTGSTATPSSQGPGAG 2366
QY 762 -----SEVSTPTT----- 770
Db 2367 TPHVSTATTPTVTSSKATPSSPGTATALPALRSTATTPTATPSSSLGTTWTRL 2426
QY 771 -----KEPTTHKS----- 779
Db 2427 SQTTPMATMSTATPSTPETVHTSTVLTTATTGATGCVATPSSPGTAHTTKVPTTT 2486
QY 780 ----- 779
Db 2487 TTGFTVPSSSPGTARTPPVWISITTTTPTSGSTVTPSSIPGTHTPVLTITTTQPVATG 2546
QY 780 -----PDEST-----PELSAETPKALEN-- 798
Db 2547 SMATPSSSTQSGTPPSLTATTATTATGSTTNPSSTPGTTPIPPELLTATTAPATSS 2606

QY 799 -----SPKEP----- 803
Db 2607 VTPSSALGTHTPPVNTTATTHGRSLSPSPSHVPTAWTSATSGTLGTHITHEPSTGTS 2666
QY 804 -----GVPTTKTTPAATKPEMT-----TTAKDKTTER----- 829
Db 2667 HTPAATTGTTTSTFALSSPHSPSTTESPPSPGTTTGHHTTATSKTTATATPSTKTRTST 2726
QY 830 ----- 829
Db 2727 LLLPSQTSAPITTVVTVTCGCEPOCAWSEWLDYSYMPGPGGDFDYSNIRAGAVCEQP 2786
QY 830 ----- 829
Db 2787 LGLECRATAQPGVPLGELQOVVECSLDFGLVCRNREQVKFCNFVEIRVFCNCHCP 2846
QY 830 -----DLRTTPETT-----TAAPKMTKETA-----TTTE 853
Db 2847 STPATSSAMPSTPGTWTWLLTTELTTTATTASTGSTATPSTPGTAPPPKVLTSPTATP 2906
QY 854 KTTESKI-----TATTQV--TSTTQDT-----TPFKITTLKTTTLAPKVTTTK--KTIIT 900
Db 2907 TATSSKATSSSPRTATLPVLSTATKSTATSVTPIPSSLTGTTGLPEQTTTPVATMS 2966
QY 901 TTEIMNKPEE-----TAKPKDRATNSKAT-----TPK 927
Db 2967 TIHPSSTPETHTSTVLTKAITTRATSTSTPSTPGTWTWLLTTELTTAATTAGTGTAT 3026
QY 928 POK----- 930
Db 3027 PSSPTGTTWLLTTELTTTATTASTGSTATLSTSTPGTWTWLLTEPSTTATVATPPGSTATAS 3086
QY 931 -----PTKAPKPTSTKPKXT--MPRVKPKPTTPTRKMT----- 963
Db 3087 STOATAGTPHVSTTATPTVTSSKATPSSSPGTATALPALRSTATTPTATSPATPSSSL 3146
QY 964 -----STMPELNPTSRAEAMLOTT-----TRNQTNPNSKLV 995
Db 3147 GTTWRLSQTTPTATMSTATPSTPETVHTSTVLTTATTATGATGCVATPSSPGTAHT 3206
QY 996 EWNPKSDAG--GAEGETHPMLLRPHVEM-----PEVTPDMXYLR 1034
Db 3207 TKVPTTTTGTATPSSSPGTALTPVWISITTTTPTTPTTSGSTVTPSSIPGTHHTAR 3266
QY 1035 VPN-----QGIIINPMLSDETNICNGKPDVGLTLRN-----GTLVAFR 1073
Db 3267 VLTSTTTTATGSMATPSSSTQT--SGTPPSLTATTATTATGSTTNPSSTPGT----- 3318
QY 1074 GHYFWMLSPSP-----PSPA 1089
Db 3319 -----TPIPPVLTSMAATTPAATSSKATSSSPRTATLPVLSTATKSTATSFTPIPS 3371
QY 1090 RRITEVMGIP-----SPIDTVEFRNCE-----GK 1114
Db 3372 STLWTTWTVPAQTTPMSTMTSIHTSSPTETHSTVLTTTATWTRATNSTATPSSILGT 3431
QY 1115 TFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQOIVAAL-----STAKYKNWP 1163
Db 3432 T-----RIITELTTTATTAAATGSTATLSSTP 3458
QY 1164 ESVYFFXEGGSIQOYIYKQEPVQKPGRRPALNVPVVGEMTOVRRRPERAIGESQHTI 1223
Db 3459 GTTWI-----LTE-----PSTIATV 3473
QY 1224 RIQYSPARLAYQDKGVHLNWKVSILWRGLPNVVTSAISLNPINRKPDPGYDYAFSKDQY 1283
Db 3474 MVPTGSTATTSSILGTAHT-----PKVVITAMATMP----- 3503
QY 1284 NIDVPSRTARAITTRSGOTL 1303
Db 3504 -----TATASTVPSSSTV 3516

RESULT 5
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphanac into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:gi1666688; PID:gi1666689; PIDN:AA8187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphanac into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 14.5%; Score 1021.4; DB 2; Length 2187;
Best Local Similarity 20.4%; Pred. No. 9.6e-14; Indels 896; Gaps 81;
Matches 392; Conservative 159; Mismatches 479; Indels 896; Gaps 81;

4 KTLPIYLL-----LLSFVIQVSSODLSCAGRCGEYSR 40
479 KNLPIALVNVGAPVSPAQAGLPRKDTTLQPLALIKESPSQASSL-----EVLSE 533
41 D-----ATCNDYCNCHMECCDFKRVCTAELSCKRCFESFE 79
534 DTVTKTTGGPAPVVRPAIGVAT-----TTL- 561
80 RGRECDCAQCKYDKC-----C 97
562 ---RADSPAVIRADSCVSPNTVQPLKRSVTDPAAPRTAKNTAPSTTSLVPLASGCG 618
98 PDYESFCRAVDKNNRT-----KKKPTPKPPV-----VDEAGSGLDNGDF 138
619 PVASSMALSQNASVSETALASPEIKSVPPDPLAEISFSPNARKVDVSHMESSGSS 678
139 KVTTPDSTTQH----- 150
679 RQGHDPASVTAKGTVVCLADSLDTSVSASKGSLSGASSPLYLEVSLPEAGLAVQGP 738
151 ---NKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEIK- 195
739 KGSNLKSLTPPSSKGAPV-PSTGAPP---SPKGAPIVPTESSISSKQVPAEILPSPQK 793
196 -----TTTNKOTSTDGKTKTSKAKETQSIKTS- 225
794 TPEVTASRLISAVQPKVDPMDSVDVTPSPKTSATAVPKDTS- -TSLKSVPAVTSLS 851
226 ---KDLAPTSKVLAP-----TPKAETTTKG 248
852 PPKAPVAPSNEATIVTEIPTSLKNALAAATPKETIATSIPIKVTSPSPKTPKSVSLKGA 911
249 PALTT-----PTPTTPKEPASTT-----PKEPTTP 275
912 PAMTSKATEIAASKDVSPQFKEVPLLOHVPTSPPKSPVSDTLTSGALTSPPKGP-P 970
276 TTIKSAPTPKE-PAPTTTKSAPTPK-----EPAPTTTKEPAPTTTKEPAPTTTK 325
971 ATLAEPTYPKSPKPAASKTTPATPSRPGVTAVPLEIPPCSKKAPKTAAPKESATSSS 1030
326 EPAP-----TTTKSAPTTKEPAPTTPKK-PAPTTPK- 357
1031 KRAPKTVANSKEIPSGVTAVPLEISLPKETSKA--TGEKSASSPKSPKTAGEKTP 1088
358 -----PAPTTTKEPTT-----TTPK- 373

Db 1089 PGGVTAVPPEISLPPKETPONATPNESLAASSQKRSPKTSVPKTPPGVGTAMPLEIPSA 1148
Qy 374 --PAPTTKEP-----APTTKEPAPTAPKKPAPTTP----- 402
Db 1149 PQKAPKTAVPKQIPTPEDAVTILAGSPLSPKASKASTAAKPEAPATPSVGVIAVSGEISPS 1208
Qy 403 -----KEPAPTTKEPAPTTTKE-----PSPTTPKE---P 429
Db 1209 PKTSKTAAPKENSATLPPKRSKPTAAKPE--TPATSEGVTAVPSEISPSPTTASKGV 1267
Qy 430 APPTTKSAPTTTKEPAPTTTTSAPTT--PREPS--PTTTKEPAPTTTKEPAPTTTKEPAP 485
Db 1268 VTLTPKGAPNALAE-SPASPKKVPKTAAPETSTTPSPQIKPVAGPKAASATPSKKTTP 1326
Qy 486 TT-----PKEPAPTTTKEPAPTTTKEPAPTKPE-PAPTTKEPAPTTPKKTTPPT 536
Db 1327 KTAVPKETSAPSEGVTAVPLEIPSPKAPKTAAPKETPAPS--PEGATTAPVQI-PPSP 1383
Qy 537 EKLAPTTPEKAPTTPEELAPTTPEEPTTTPKEPAPT--TPKAAAP----- 581
Db 1384 RGSKKAGSKETPTTP-----SPEGVTAAPLEIPISKKTSKMASPKETLVTSPSKKLS 1437
Qy 582 -----NTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTPKGTAPTTLKEPAPTTPKP 635
Db 1438 QTVGKETSLEGATAVPLEIPSHKKAQKTVDPKQVPLTPSPKDPATTLAE-SPSSPK- 1495
Qy 636 APKELAP-----TTTKEPSTTSKDPAPTTTKEPAPTTPKGTAPTTLKEPAPTTPKPA 682
Db 1496 APKTAAPSERVTTVTP-----EKPA-TPQKASGTTASKVPVPAETQEVAVSSRRTPTVP 1549
Qy 683 TTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTPKGTSTTSOKPAPTTPKGTAPTTPKEPA 742
Db 1550 AVPPVKNSSHKTSKTIELKEAPATLPPSPKSPKIPSSKX---APRTSAP----- 1598
Qy 743 PTPKAPAPTTPETPTTSEVSTPTTKEPPTTIHKSDESTPE--LSAETPKALENSP 800
Db 1599 -----KEFPASPSIKPVTTSLAQ-----APPSLOKAPSTTIPKENLAA---PAVLVSS 1645
Qy 801 KEPGVPTTKTAAATKPEMTTAKDKTTERDLTTP-----ETTTA-----APK 843
Db 1646 KSPAAP-ARASASLSP---ATAAPQTAPKEATTIPSCKAAATETPIETSTAPSLEGAPK 1701
Qy 844 MKETATTTTEKTESKITATTTQVSTTTTQDTPPKITTL-----KTTILAPKVTTTKTI 899
Db 1702 ETSEISV-----SKVLMSPPKASSSKASTLPATLPSLKEASVLSPTATSGK-- 1752
Qy 900 TTEIMNKEETAKPKDRATNSKATTPKPKPTKAPK----- 937
Db 1753 -----DSHISVSDACSTGTTT--PQASEKLPSKKGPTAFTMLAAPAPESALAIT 1801
Qy 938 -----PTSTKPKKTMPR--VRPKTTTTPKMTSTWPELNPTSR 974
Db 1802 APIQKSPGANSNSASPKCPDPSKKOTKGLPSAVALAPQTVPVKE-----DTSK 1851
Qy 975 TAEAMLQTTTRENQ----- 988
Db 1852 ALETLLVSPAKGDCCLHSPKGPVGSQVATPLAAFTSDKVPPEAVSASVAPKPAASLTL 1911
Qy 989 --TPNSKLVENVPKSEDAGGAETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPML 1046
Db 1912 APSVPAPLPPKQPLLESAPGSVLESKIL-----PVPAAEDEL-PLI 1953
Qy 1047 SDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSP-----PSFARITEVWGLP 1099
Db 1954 PPEA-VSGE-----PFQDILNMPAPKEA-----GTP 1980
Qy 1100 SPIDVTFRNCNCEKGTFFFKDSQYWRFTNDIKDAGYKPIFKGFG----- 1145
Db 1981 APAPSA-----KQVLLKNNKSGSTESDESVPBL 2010
Qy 1146 -----LTGOIYAALSTAKYKNWESVYFFKRGSGSIQYIYKQEPVQKCPGRPAL 1195
Db 2011 BEQDSTQTATQAAQAAAAE-----IDEEPVSQAKQSR--- 2043

[illegible]

QY 380 -----BPAP----- 383
Db 1583 EDINECAKAPCSLNCANVMNCTFSCSKQGYRGDFMCTDINECDRHPCHPAECT 1642
QY 384 -----TTPKE----- 388
Db 1643 NLEGSFKCECHSGFEGDGKIKCTNPLERSCEBDEVEKFCGRVDHVSLSVRIYNGSLSSVCE 1702
QY 389 -----PA----- 390
Db 1703 CEPGRFEKESNCVDIDECEBSRNNCDPASAVCVNTEGYSRCECARGYEGEGVCTDID 1762
QY 391 -----PTAPK----- 395
Db 1763 ECDRGMAGCDMSAMCINRMGSCGCKMAGYTGDCATCIKEEPKSKDTACTDEWSLCE 1822
QY 396 -----KPAPTTKEPAP----- 407
Db 1823 LEKKQCTVDEEVEPQCGACLPFGHHPINGTCQSLQISGLCAQKNDCKNKAECIDIHDPDSHF 1882
QY 408 -----TTP 410
Db 1883 CSCPDGFIGDGMICDDVDECNMAGMCDENTKENTIGSFNCVCLGFKKVDKCVVDEK 1942
QY 411 KEP-----APTITKEPSPTTKEPAPTITTKSAPTITKEPAPTITTKSAPTITKE 458
Db 1943 KQPNREKIEIDENSSSSNGQEKPTKGI VSSTSATSSSESTTAEBHVITISSTSTKD 2002
QY 459 PSTITKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKKPAPTAPKEPAPT 518
Db 2003 MTSSKSPENVMTSSSEFVSTSSSKSTTASSETTVSSSTPSESS-----SSEAPLTSSPA 2055
QY 519 TPKETAPTTPKLTPTTPE-----KLAPTPE-----KPAPTITPEELAPTTPPEP 563
Db 2056 TTTEV-----ITESSVKSTTPKEESSSEITVKLSSKSPETVSSSVKSSPSTPS-----TTSQSV 2109
QY 564 TPTTPEEAPTTPKAAAPTTPKEP-----APTTPKEPAPTTPKEPAPTTPKET----- 611
Db 2110 TSTVPETSKSTVLSSEAPVTSPTSEVHTSSSETKPSLSASSITGDNSTTPSTSSLASVK 2169
QY 612 APTTPKGT-----APTILKEPAPTTPKKAPKELAPT-----TTKEP----- 648
Db 2170 STSAPEGTSASVAPVKLSLSPDV-SQPSTKTFDATESSTVQASSETSGTSVKSTSEPS 2228
QY 649 -----TSTTSDKAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG----- 687
Db 2229 HVTKLSITSSNFSSVVPVTPSKSTPTVPPEST-EQPTSTTPSGQSLTPMNSNSEVLTTSE 2286
QY 688 -----TAPTTLKEPAPT-TPKKAPKEL-----APTTPKGT-----ST 720
Db 2287 PHVLSSLSLSPDVSQSTTENNLSSESTVETPKTSSEVLSNSEPTTEAPTTLSPDILST 2346
QY 721 T-----SKP-----APTTP----- 730
Db 2347 TTNNLQSSTVSTEDRSEISSENSEKPSAPBLVTSVTHVASSPDVPTESSEPDPLTG 2406
QY 731 -----KETAPTTP-----KEPAPTTPKKP-----APTPE----- 755
Db 2407 SSTEINPEASSKQTISSPTPTTTTASEPTKSTMSDPLSTTSNVLSESTTPESSSKS 2466
QY 756 -----TTPPTTSEVSTPTTTPKEPTT----- 775
Db 2467 PVSSSTEGISVVSTEFKSVPESTISSVLEEDLTKTTPSPILEETTTTASETSEPLETDSL 2526
QY 776 -----LHK--SPDESTPELSAPTTPKALENSP-KEPG-----VPTT----- 808
Db 2527 TVSVRIHELTTSENVPKSESTTTSSBSKPSQEPAGILTSVVVPTSSVSLITASIE 2586
QY 809 -----KTPAAKPEM-----TTTAKDKTTERDLRTTPTTTTAAPOKMTKETA 849
Db 2587 AITSNTPKQGRTPITTPGKSLVSKSTTSPSTVTSSEPSSESTKRITTVSTVSTTPTTEBT 2646

QY 850 TT-----TEKTTESKITATTQVTS-----TTQDTTPF 878
Db 2647 TSLSLILTAAPSKPTTESTTESSEAPTTPAKTSETKPNVSVSTSRKSTENVETSSQSGSL 2706
QY 879 KITTLKTT-----TLAPKVTTTKTITTT----- 902
Db 2707 ESSIMSTSEPEINAPAVTVSSEASTTLEENSTSSPTSSEASVKLSLFPESITSEA 2766
QY 903 -----EIMNKPPE----- 910
Db 2767 VTVSSRAPAEITWSSSEHREISTVSSEPSSEPELPTTVSPNVVTVASSIFSEEPILSSVT 2826
QY 911 -----TAPKD-----RATNSKA-----TTPKPKQ 930
Db 2827 SSSTPRVRLTGTGPDLLIVSVTVPSHGNNRRQNTASSVPSNSTPILPSESITTPQPPP 2886
QY 931 PTKAPKPTSTTKPKMTPVRKPK-----TTPETP-----RKWTSTMPELNP 971
Db 2887 TTTTAKPATTSKGRPPSIQPPAEMFTTPAPPPNSGGYGEETNOEEQVTS----- 2940
QY 972 TSRIAEAMLQTTT----- 984
Db 2941 TTTEAPSLCSTVTCCHSLATCEQSTGVCI CRDGFIDGTTACSKKSTADCI SLPSLCAK 3000
QY 985 -RPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFPNQGIIN 1043
Db 3001 KCDNSTRSCEC-----DAG-----YI-----GDGVCS 3023
QY 1044 PMLSD-----ETNICNGKPV-----DGLTLR-----NGT 1068
Db 3024 PHQDCVLRDNLCSPAVQNNRCQCLPGFTGDKVCSIHERRASNCQCDANAHCVGT 3083
QY 1069 LVAFRGHYFWMLSPPSPSPARRITEVWG-----IPSPIDTV-FI-----R 1108
Db 3084 TCKCNPGYF-----GNGLCCVPDPLDCVHTFGICHFNVAVCNSESQ 3124
QY 1109 CNCEGKTFEFKSOYWRFTNDIKDGYPKPIFKGFGG-----LTGQIVAAALSTA 1157
Db 3125 CQCS-----SGFSNGVSGCFPPQKSCRTDKSVCA----- 3152
QY 1158 KYKNWPESVYVFPKGGSIQYIYKQEPVOKCP-----GRPEA-----LN 1196
Db 3153 -----KNAICLPTGSCICRHGFKGDPFYKCSLSVSKEPANQODLSDVSSCVTPCDSSTQ 3206
QY 1197 YPVYGMTQVRRRRFER--AIGPSQT----- 1220
Db 3207 LCISGEC--ICKSGFRNSSLGSETCADI DECAEKSHKCDRAVATCRNTFGSHVCTCPDG 3264
QY 1221 -----HTIRIQSPARLAYQDKGVLHNEVKVSLWR--GL----- 1253
Db 3265 HVGDGITCVPHV-----NOGKLSVYCEADGMTVLGLNETADPEGKI 3305
QY 1254 -----PNVVTSAISLPNIRKP----- 1269
Db 3306 FVKGGAENPYCKSPSSLLNSRKPVFKVFOHCDVQLLDNHTMASTVVVQKHAMFLTNK 3365
QY 1270 -DGYDY-----YAFSKDOYVNIID----- 1286
Db 3366 ADSYDLRCQYPIGSRVASHVNVSELAITSTLTTEKNSTLAPICRLSVSNDQHSISSAMV 3425
QY 1287 -----VPS-----RTAATITRSGQ 1301
Db 3426 GDTLKALEVTPESENFGLPRNCFAINTESGE 3457

RESULT 10

T29757

Protein UNC-89 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C:Accession: T29757

R:Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of C. elegans cosmid C09D1.
 A;Reference number: Z20679
 A;Accession: T29757
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-6642 <DUG>
 A;Cross-references: EMBL:AF003131; PIDN:AAH54132.1; GSPDB:GN00019; CBSP:unc-89
 A;Experimental source: strain Bristol N2; clone C09D1
 C;Genetics:
 A;Gene: CESP:unc-89
 A;Map position: 1
 A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
 /3; 5917/1; 6021/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1
 Query Match 13.4%; Score 945; DB 2; Length 6642;
 Best Local Similarity 12.2%; Pred. No. 4.9e-11;
 Matches 480; Conservative 190; Mismatches 470; Indels 2780; Gaps 123;
 34 CCEGYSRDATCNG-----DYNCOH----- 52
 659 CKEGVIDFCEVGEWPELVWDDQPLRSHDFRLQYDQGTAKLEIRDAQPDGTGY 718
 53 -----YMECCPD-----FKRVCFAE-- 67
 719 TVKIQNEFGSIESKAEFLVQADFDKNHVAPEFOATIEYVEC--DEGEVRFKSVITGDPN 776
 68 -----LSCKGRCPESFERGECDC 86
 777 PEIIFWINGKLSEKVKFISEGICILTIKDVTRHFDGMVTCOS-----NRLGSASC 831
 87 DAQCKYDKCCPYESFCAEVKDNKNRTKKTPKPPVVDE----- 128
 832 D-----GRUKVRVPPAPPTFNKLEBKTVOEKSIVVFEVD 866
 129 -----AGSLONGD-----FKVTTPTDSTQHN----- 151
 867 VSGWPEPTLTLCGKELNGBEGVEIVGHDFYRISIPNTSMDKHGEIVAKAQNEHGT 926
 152 ----- 151
 927 AESRRLTVQEBEERSAPTFLKIDQTVKTGEFAVFETTVRGNPNPVTWFINGHKM 986
 152 ----- 151
 987 DQSGPGVKIEAHNDHKLITDSQAYAGTVLCRAENAVGRPETKARLVVLAPKQKPPKF 1046
 152 -----KVSTSPKLTIA----- 162
 1047 VEILVDKTEVDNTVVFVRVEGEPKPTVTWYLGEBELKQSDRVEIREFDGSIKISIKNI 1106
 163 -----KPINRPSLPP----- 173
 1107 KIEDAGEIRAVATNSEGSDETKAKLTQCKKFPFAPFDRLRPSLTVKEGSAVFSAHAFGI 1166
 174 ----- 173
 1167 PLPTYEWSVNGRKVRDQOEGARVTRDESTVDGASILIDITATYYSEVNNHLTISVAENTL 1226
 174 -----NSDTSKE-----TSLTVNKE 189
 1227 GAETGAQLTTEPKKESVWVEKQDLSSEVQKEIAQQVKEASPEATTITMETSLTSTKT 1286
 189 TTVEKKTATTT-----NKQSTDGKKTTSKAKETQSIKTSKADLAPTS 232
 1287 TTMTTTEVTSTVGGTVTKESESATTVIGGSGGVTEGSIKSVSVKIEVVKTSQTDV 1346
 200 ----- 199
 1347 REGTPKRVSPAEELPKVEIDSDRKKKSPDPKKEK-----SPEKTEEPASPTK 1398
 233 KVLAKPTPKAETTTKGPALTTTPKEPTTP-----KPASTTPKPT-----PTTIKS 280
 1399 KT-----GEEVGPKEKSPASPTKSKSPAAEVRVKSPTTKKESFSSPTKKEK 1445

281 APTTP-----KBPAPTTTKSAPTTPKEPAPTTTKBPAPTP-----KBPAPTTT 324
 1446 SPSSPTKKTGEVKEKSP-----PKSP---TKKESPEKPDVKSVPVKEXSPDAT 1493
 325 -----KBPAPTTTKSAPTTP----- 339
 1494 NIVEVSSETTIKTETTTMTTMTHESESRSTSVKKEKTPKVDKPKSKTKDKSPEKSI 1553
 340 -----KBPAP--TTPKKPAPTTPKBPAPTTTPKEPTTPKE--PAPTTKEPAP----- 383
 1554 TBEIKSPVKESPKBEKSPASPTKKEKSPKBPASPTKKSNEVKSPTKKEKSPKSVV 1613
 384 ---TTPKEPAP--TAPKKPAPTTTPKEPAPTTTKBPAPTTTKBPAPTTTKSAPT 439
 1614 BELKSPKESPEKADDKSPKTKK-----KSPEKSATEDVKSPTKKESPEKVEKPT 1667
 440 TTKBPAPTTTKSAPT-----TPKEPSTTTTKE--PAPTTTPKEPAP-----TTP 480
 1668 -----SPTKKESSPTKKTDDVKSPTKKESPTKKEKSPQTVEEKPASPTKKEKSPKSVVEVKSP 1722
 481 KKPAPTTTKE--PAPTTTPKEPAP-----TTTTK---PAPTAPKEPAPTTTPKETAPTT 527
 1723 KKSPEKABEEKSPKTKKESPEKSAAEVKSPTKKESPEKSAEEKPKSPTKKESSPV- 1781
 528 PKKL-----TPTTPEKLAPTTPEKPAPTTTPBELAP--TTPPEPTTTPPEEPAPTTPKAA 579
 1782 ---KMAADDEVKSPTKKESPEKVEEKFASPTKKEKTPKSAABELKSPTKKESPSPTKK 1839
 580 APNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTLKEPAPTTPKKAPK- 638
 1840 TGDSEKESPEKPEE---KPKSP---TPKSPGSPK-----KKKSKSPEAKPPAPKL 1887
 639 -----ELA----- 641
 1888 TRDLKLQTNKTDLAHFVVVVEHATECKWFLDGEKITTAGQVTVSKDQOPEFRCSDITTM 1947
 642 -----PTTTKEPTST----- 651
 1948 FSGSTVSVASNAAGSVETKELKVLTPKETKKEPFTDKLRDWEVTKGDTVQMDVIALH 2007
 652 -----TSDKPAPTT 660
 2008 SPLKYQWQNGLLDGKNGVTIKNEBKNKSLIIPNAQDSKGITVBSASNEVGSSSQAOL 2067
 661 --PKGTAPTTPEKPAPTTKE-----PAPT----- 683
 2068 VNPSTTPIVVDGPKSVTIKETETAEFKATISGFPAPTVKWTINEKIVEBSRITTIKTE 2127
 684 -----TPK----- 686
 2128 DVYTLKISNAKIBQGTGVKVTQAQNSAGQDSKQADLKVEPNVKAPFKKSQLTDKVADEGP 2187
 687 -----GTAPTTLKEPAP----- 698
 2188 LRWNLELDGSPGTEVSWLLNGOPLTKSDTVQVVDHGDGTGTYHTVIAEAKPEMSGTLTAKA 2247
 699 -----TTPK-----KPAPK----- 707
 2248 KNAAGECETSAKVTVNGGNKKPEFVQAPQNHETLEESVKESAIIVTKGPMNVWYLNKK 2307
 708 ----- 710
 2308 KLIQSEEVKVVYHETGKTSIRIQKPLMEHNGTIRVEAENVSGKVQATAQLKVDKKEVTP 2367
 711 PTTT-----KGPTSSTSDKP-----A 726
 2368 KFTTMDRQVKEGEDVKFTANVEGPEPSVAVNTLNGEPVSKHNITVTDKGBHTTIEIS 2427
 727 PTTPEKT-----APTPK----- 739
 2428 AVTPEQAGELSCBATNPVSGKRDVQLAVKKVGADPTFAKNLEDRLITEGELTLMDAKLN 2487

QY 740 ----- 739
 Db 2488 IVKPKPTIWLKDGVEITSDGHYKIVEEDGSLKLSILOLKLKEDKGRITIKAESFGVAE 2547
 QY 740 ----- 745
 Db 2548 CSASLGWVKGRMAKPAFQSDIAPINLTGDTLEKLLITGDTFFVKWIGTQVCATE 2607
 QY 746 ----- 756
 Db 2608 DTEISNANGVYTKIHGVTAADMTGKIKCVAYNKAAGEVSTEGPLKLVAPIVPEFETSLCDA 2667
 QY 757 ----- 775
 Db 2668 TCREGDTLKLRAVLGEPFVWVWYNGKLBESQNIKHSEKGTIVTITKIDTCYSGQ 2727
 QY 776 ----- 778
 Db 2728 WVCEAINEXKATSEATLVLPRGEPDFLEMLSNVRARTGTQVHKVWFTGDPKPSLTW 2787
 QY 779 ----- 781
 Db 2788 YINKEILNSDLTIYDDKSTLTINSFNPDPVHVGEILICKAENDAGEVSTANMITYS 2847
 QY 782 ----- 801
 Db 2848 DMFSESESAQAEEFVGDDLTDEBSLEEMHRTPTPM--APKFTIKIDKAKKGHSA 2904
 QY 802 ----- 809
 Db 2905 VFECVVPDTGVCCKWLDGKEILIRIARVQRTGPEGHITQELVLDNVTPEDAGKYTC 2964
 QY 810 ----- 836
 Db 2965 IVENTAGKTCATLVIESLEKSKKAPFIVALQDKTTTSEKXVLECKVIGBPKK 3024
 QY 837 ----- 887
 Db 3025 VSWLHDNVREKNPSEKTIQESITVESGVERVTISSE-----LSHQKYTC 3075
 QY 888 ----- 910
 Db 3076 IAEENTGSKTEAFLTVQGEAPVFTKELQNKELSIGELVLSCSVKGSQPHWDFYFSE 3135
 QY 911 ----- 932
 Db 3136 TTKVETKISSRIAEHDQTNHWRMVISQITKEDIVSYKAIATNSIGTATSTSKITK 3195
 QY 933 ----- 949
 Db 3196 VEAPVPEQGLKTSVKEKEEIKMEVKVGGSPDVEWFKDDKPVSEGNHMKKNPETGVF 3255
 QY 950 ----- 975
 Db 3256 TLVWQAAATTDAGKYTAKASNPAAGTAESSAEVTSLEKPTFVRELVTTEVKINET-- 3312
 QY 976 ----- 995
 Db 3313 --ATLSVTGVDPDSVWELKDGQPVQDTSSHVIAKVEGSGSYSITIKDARLEDSGKYAC 3370
 QY 996 ----- 1027
 Db 3371 RATNP-----AGEAKTEANFAVKNLVPPEFVEKLSFLEVKESITLSVKVVGTPPEPSVE 3426
 QY 1028 ----- 1041
 Db 3427 WFKDDTPISDNVHVIQQTAVGSPSLTINDARQDVGIYSCRNARNEAGEALTANFGII 3486
 QY 1042 ----- 1059
 Db 3487 RDSIPPETQKLRPLREVEQETLIDLKVTIGTVPVPEVWFKDD-----KPINIDNSHIP 3540
 QY 1060 ----- 1081
 Db 3541 AKDEGSHHTL--TTQARGEDGVVYTCATNEAGEAKTTANMAVOEEIEAPLFVQGLK 3597

QY 1082 PF--SPSPARRITEVWGISPI----- 1102
 Db 3598 PYVEQCKPAELVVRVEGPEPEVKFKDGVPIADINQHVIEKKGENGSHTLVIKOTNNA 3657
 QY 1103 ----- 1109
 Db 3658 DFGKTCQATNKAGKDETVCGLKIPKYSFEKQTAEEVKPLFIPLKETAPEVGDTVVLEC 3717
 QY 1110 NCEGKTF--FEKDSQYWRP----- 1126
 Db 3718 KVNKESHPQIKFPKNDQPVIEIGHQMLEVLEDGNIKLTIQNAKKEDGVAYCEAVNVAGK 3777
 QY 1127 --TN-DIK-----DAGYKPKPIF----- 1140
 Db 3778 ANTADLKIQAPAAKVEHVTDSESGLEIEIQFETVGDTSASKTDTGRGAPEFVELLSCT 3837
 QY 1141 ----- 1147
 Db 3838 VTEKQAILKCKVKGPRPKIKWTKEGKEVMSARVRAEHKDDGTLTLTFDNVTQADAGE 3897
 QY 1148 ----- 1163
 Db 3898 YRCEAENYGSAAWTEGPIIIVTLEGAPKIDGEAPDFLOPVKPAVTVGTAVLEKISGKP 3957
 QY 1164 E-SVPFFKRGGSIQ-----QY----- 1178
 Db 3958 KFSVKWYKNGEELKPSDRVKIENLDDGTQRLVTNNAKLDMEYRCEASNEFGDWSVDT 4017
 QY 1179 IVKQBPVQKCPRRPALNYPVYGMTQVRRRPERA----- 1214
 Db 4018 LTVKEPAQVAPG-----FFKELSAIQVKETETAKFECKVSGTKPDVKWPKDGTPLKE 4069
 QY 1215 -----IGPSQTH--TIRIQYSPARLAYQKGVLUHNEVKVSI----- 1248
 Db 4070 DKRVHFESTDGTQRLVIEDSKTDDQGNRYIEVS-----NDAGVANSKVPLTVVPSETL 4123
 QY 1249 -LMWGLPNV-----VTSA----- 1260
 Db 4124 KIKKGLTDVNVVQGTKILLSVEVEGPKTVKWKYKTEIVTSSQTKIVQVTESEYKLEIE 4183
 QY 1261 -----ISUPNINK----- 1268
 Db 4184 SAEMSDTGAYRVVLSTDFSVESSATVTVTKAAEKISLSPFKKGLADQSVPGTPLVLEV 4243
 QY 1269 -----PD----- 1270
 Db 4244 EIEGPKQVKVYKNGDEIKDGKVEDLGNKVRILTIPDFQEKDVGYSVTAANEAGEIESK 4303
 QY 1271 -----GDYVYAFSK-----DQY 1282
 Db 4304 AKVNVSAKPEIVSGLVPTTVKGETATFNVKVGPKVGVKWKYKNGKEIPDAKTONGDGS 4363
 QY 1283 YNIDVP-----SRTARAI-----TTRS 1299
 Db 4364 YSLEIPNAQVEDAADYKVVWSNDAGDADSAALTVKLADDDGKDKVKPEIVSGLITFTVKQ 4423
 QY 1300 GQTLN--KV-----WY 1308
 Db 4424 GETATFNVKVGPKVQVKWY 4443

RESULT 11

S49915

extensin-like protein - maize

C:Species: Zea mays (maize)

C>Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004

C:Accession: S49915

R:Rubinstein, A.L.: Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Pex genes: pollen-specific genes with extensin-like domains.

A;Reference number: S49915

A;Accession: S49915

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1188 <RUB>

A;Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g6001

Query Match 12.7%; Score 900.4; DB 2; Length 1188;

Best Local Similarity 18.9%; Pred. No. 1e-11;

Matches 290; Conservative 78; Mismatches 407; Indels 756; Gaps 54;

QY 10 LLLL---LSVFVIOVSSODLSSCA-----GRCGR-----36
DB 17 LLLLAACLSACSQVAVTSAEAYIAHRLQLLAMKEAGGAGDLPADFEFDRRVGAANFPN 76
QY 37 -----GYSRDATCN-----CDYNQHYMECCPDFKRVCTAEL 68
DB 77 PRLRAYIALQAWHRAFYSDPKGY---TANWVGEDVCKYNGVICTEALDDPKITVWAGI 132
QY 69 SCKGR-----73
DB 133 DLNGADIAAGYLPPELGLLTLDAFFHINTRFCGIIIPKMSRLSLHLEFDDVSNRNFVGVPP 192
QY 74 --CFPS-----FERGREGDCDA-----88
DB 193 YVCEWVSLKYLDLREDFEGELPALFDK---DLDAIFVNTNRVGVGIPENLGNSTAS 248
QY 89 -----QC-----KKYDKCCP-----98
DB 249 VIVFANNAFVGCIPIKSGRMVKTLDIEIFLNKLDGCLPLEMGLLVNTTVIDVSGNMLVG 308
QY 99 -----DYESFC-----104
DB 309 TIPEQLSNIACLEQLDVRNFTGIVHESICELPALVNFSPAPNFPNSBAAVCMPSDKAL 368
QY 105 -----AEVKONKNRKTKKTPPKPVVDEAGSLDNGD 137
DB 369 VNLDRDNLGALRPAQKALQCAPVLARPVDCSHKVCAGYTP-----GGG-----415
QY 138 FKVVTPDTSTTOHNVKSTPKITAKPINRPSLPNSDTSKTSLSLVNKETTVETKETT 197
DB 416 -----PPSSPVFGKPAASAPM-----PSPHTPP--DVSPEP-----444
QY 198 TTNKQTSIDCKEKTSAKTSQIEKTSAKDLAPTSKVLAKTPKAEITTKGPAITTPKEP 257
DB 445 -----LPEPSVPAPAPMPMPTHSPPADDYVVP 473
QY 258 TPPTTP-KEPASTTP-----KEPTPTTKSAPT-----TPKEPAPTTKSAPT-TPK 301
DB 474 TTPVCKSPATSPSPQVOPPAASTPPSLVKLSPPQAPVGSPPPPVKITSPAPIGSPS 533
QY 302 EPAPTTTKEPAPTTKKEPAPTTKKEPAPTTKSAPTTP-KEPAPTTKKEPAPTTKKEPAP 360
DB 534 PPPPVSVWSPPPPVKSPPPAPVGSPPPEKSPPPAPVASPPPPVKSPPPTPLVASPPP 593
QY 361 TTPKEPTTPKKEPAPTTKKEPAPTTTP-KEPAPTPKKEPAPTTKKEPAPTTKKEPAPTTTK 419
DB 594 FVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVASPPPPPMKSPPPPTPVSSPPPEK 653
QY 420 EPSF-----TTPKEPAPTTTKSAPTTPKKEPAPTTTKSAPTTPKKEPAPTTTKKEPAP-PTTP 472
DB 654 SPPPPPAKSTPPPEEYPT-----PPTSVKSSPPPEKSLUPPTLTPSPQEKPTPPSTP 708
QY 473 KEAPTTKKEPAPTTKKEPAP-----PTTPKE-----498
DB 709 SKD-PSSPEKPSD--PKEPVSSPQTPKSPPPAPVSSPPPTPVSSPPPALAPVSSPPSVK 765
QY 499 -----BAPTTTKKEPAPTPKEPAPTT-----TPKETAPTTPKXLTTPPEKLAPTTPKEP 547
DB 766 SSPPPALSSPPAPQVKSPPPVQVSSPPPPAPKSSPPLAP-----VSSPPQVEKTSPP-PP 820
QY 548 APPTPEELAP-----TTPPEPTTTPPEPAPTTTKAAAPNTPKPEPAP 589

DB 821 APLSPPLAPKSSPPHVVVSSPPPVVKSPPAPVSSPPLTPKPDASPAHVSSPPEVVKP 880
QY 590 TTPKEPAPTTKKEPAPTTKKEPAPTTTP-----KGTAPTTLKKEPAPTTTPKPKAPKELA 641
DB 881 STP--PAPTWTISP--PSEPKSSPPTPVSLPPPIVKSPPPPAMVSSPPMTKSSPP-----933
QY 642 PTTTKEPTSTTSKDPAPTTTPKGTATTPKEPAPTTPKKE-PAPTTPKGTATTPTLKKEPAPTT 700
DB 934 FVVVSSPPPTVKSSP-PPAPVSSPPATPKSSPPAPVNLPPPEVKSSPPPTPVSSPPAP 992
QY 701 PKKPAKELAPTTTKGTPTSTTSKDPAPTTPKETATTPKKEPAPTTKKEPAPTTTPPTT 760
DB 993 KSSPPP---APMSSPPPEVKS--PPAPVSSPPPVKSPPPAPVSSPPPVKSPPP 1047
QY 761 TSEVSTPTTKKEPTTIHKSPDESITELSAEPTPKALENSPKPEGVPPTTKTAATKPEMTT 820
DB 1048 APVSSPPPPPVKSSPPP--PAPISSPPPVKSPPPAPVSSPP-----PPVKSPPPPAP-----1097
QY 821 TAKDKTTERDLRTTPTTTAAPKMTKATATTTKTESKITATTTQVTSITTTQDTTFFKI 880
DB 1098 -----1097
QY 881 TTLKTTTLAPKVTTTTKXITITTEIMNKBEETAKPKDRATNSKATTPKPKOKPTKAPKKPIS 940
DB 1098 -----VSSPPP--PIKSPPPAP 1113
QY 941 TKKPKTMRVRKPTTTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNK 1000
DB 1114 VSSPPAP--VKPPSLPPAPVSSPPPVVTPAP-----PK 1146
QY 1001 SEDAGAGETPHMLLRPHVMEPTPDMDYLRVFNQGIILINPMLSDETNICNGKXPVDG 1060
DB 1147 KEQS-----LP-----1153
QY 1061 LTTLANGTLVAPRGHYFWMLSPPSPSPARRITEWGPISPIDTFTTRCNCEGKTFKKD 1120
DB 1154 -----PPAESQ-----PPP-----SF-----1164
QY 1121 SQWRTNDI-----KDAGYKPKPKGF 1143
DB 1165 -----NDIILPPIMANKYASPPPPQFGY 1188

RESULT 12

T42629

N;Alternate names: flexilin

C;Species: Bos primigenius taurus (cattle)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004

C;Accession: T42629

R;Elefteriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.

J. Biol. Chem. 272, 22866-22874, 1997

A;Title: Characterization of the bovine tenascin-X.

A;Reference number: Z22180; MUID:97426436; PMID:9278449

A;Accession: T42629

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-4135 <ELS>

A;Cross-references: UNIPROT:O18977; EMBL:Y11915; NID:g2462978; PIDN:CAA72671.1; PID:g2464

C;Genetics:

A;Gene: TN-X

C;Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type

C;Keywords: extracellular matrix; glycoprotein; heptad repeat

Query Match 12.7%; Score 895.9; DB 2; Length 4135;

Best Local Similarity 12.8%; Pred. No. 2e-10;

Matches 457; Conservative 181; Mismatches 465; Indels 2651; Gaps 126;

QY 23 SSODLSS--CAGRC-GEYSRDTATCNCDY-----NCQHYMEC-----56

DB 518 AGEDCCSRRCPCDRCGRGCDGVCSDYGEGEDCGKRCPRGCGGCLGRCVCD 577

QY 57 -----CP-----DFK-----RVCTAELSKGRCFESFERG 81


```
QY 741 -----PAPTPPKPAPTTPT 756
Db 2765 EVTIGLEPRGRKYNMNLGLHSGQRPVSVGVTDPOEVVEETPSTPEAPEPEE 2824
QY 757 P----- 757
Db 2825 PLLGLDVTGSSPDSLSSLSWTVPOGHFDSFTIQYKGRDRPQVVRVGGSETEVTIGLEPR 2884
QY 758 -----PPTTSEVSTP-----TTTKEPTTIHKSPDESTPELSAEP----- 791
Db 2885 HXYKNMNLGLHSGRRVGPVSVTVGTVAPDYDAMTITQTPST-----SVPEPTPKRLGVEV 2937
QY 792 -----TPKAL-----ENSPK-----EP----- 803
Db 2938 TVTDATPDSLSSLSWTVPOGHFDFHFLVQYKNGQPKAVRVPDGEDGVITISGLEPHKYM 2997
QY 804 -----GVPTT-----KTPAAKPEMTTAKDKTKTERDLRTTPTTTAAPKM 844
Db 2998 NLYGFHDHORVGPVSVIGVTTAAEETPTKME-----ETPSPT 3037
QY 845 TKETATTTKTTES-----KITAT----- 863
Db 3038 MEETPSPTPEAPPEPEELLGELTGTGSSPDSLSSLSWTVPOGHFDSFTVQYKGRDGP 3097
QY 864 -----TTQVT----- 868
Db 3098 QVVRVGGSETEVTVGLEPRKYMHLGLHGRRVGPASTVGVTSASLTTERPLAPRLGE 3157
QY 869 -----STTQDT-----PF-----KITLTKTTLAP----- 890
Db 3158 LAVAVTSDTARLSWTVQGFDFSLVQYKDVQGPQVAVAGDLREVTVSSLPAGRYK 3217
QY 891 -----KVTTTKTITTTIEMNKPEETAKP----- 914
Db 3218 FLFLGLDERKHGFPVSADAKTLPTD-----KPAPRLGELTVDTPGSGLSWT 3267
QY 915 -----KDR-----A 918
Db 3268 PEGFDFSMVQYKDRDQRPVPAADQREVTVPGLEPNKRYKFLYGLVGRKRLGPISA 3327
QY 919 TNSKATTPKQKP----- 931
Db 3328 EGSTAPLEKERQPPRLGELTVDDETNSLRLSWTVAGRFDFSVVQYRGTDQGRMPV 3387
QY 932 -----TKAPKPK 938
Db 3388 AADQREFTVGLRPGRYKFLYGLLGGQRLGPASVLGMTAPEBDTPAPWHAATEAPKPP 3447
QY 939 ----- 938
Db 3448 EGPRGLGLAVRDSFDSLRLSWSVVGPFDFSVVQYQDTDQGPQALLVGGDQNKVLVSL 3507
QY 939 ----- 938
Db 3508 EPSTSYEFFLYGLHEGRKLGFPVSAEGTGPVAGQTPGEGPRLSHLSVTDVTTSSLRLN 3567
QY 939 -----TSTKPKTMRPRVRKPKTTTTPRKMTSTWPELNP----- 971
Db 3568 WEAPPEAFDSFLLRFGVPSPSTLPQLRPLQLRELTPVGTTR-SAVLRDLHPGLTYLTLL 3626
QY 972 -----TSRAEAML-----QITTRPNQTPNSKLVE---VNPKE 1002
Db 3627 YLGRGPHKADSIQGTARTLSPVLSPRDLQSEIRETSARVSWTPTSPTRSDGFKVSVQLA 3686
QY 1003 DAG-----GAGETPHMLLRPHVFMPEVPTDMDYLPRVPNQGIIN 1043
Db 3687 DGBEPQSVQVQDRTQKLEGLIPGAYEVTVVSVRG---FESEPLTGLTITVPD-----G 3738
QY 1044 PM-----LSDETINCNK-----PVGLTTLR 1065
Db 3739 PTHLRALNLTDESALLHWKPPQTPVDYDVKVTAFGAPFSLOASAPGSAVDYPLQGLVTH 3798
```

```
QY 1066 N--GTLVAFRGRHYF-----WMLSPSPSP----- 1086
Db 3799 NYTATLRLGLRGNFTSPASITFTTGLEAPQDLEAKVETPTALLTWTAPESPTGYLLSF 3858
QY 1087 -SPARRITEV-----WG-----IPSP 1101
Db 3859 NTPGGQTQEIILLPGGVTSQHLRGLPFPSTYTWLRAMWGDSTPPVSVSTFTTGGRIIPFP 3918
QY 1102 ID-----TVFTRCNCE-----GKTFPFKQSQ 1122
Db 3919 RDCGEMONGVSTSTFTTIFLNGRERPLNVFCMDTGGGWLVFQRRMDGKTDWFRD-- 3976
QY 1123 YWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNW--PESVYFKRGSGSQIYIY 1180
Db 3977 -W-----EDYAH-----GFGNISGEF-----WLGNEALHSLTKAGD----- 4006
QY 1181 KOEPVQKCPRRPALNYPVYGVEMTQVRRRPFERRAIGPSQTHIRIQYSPARLAYOKGYVL 1240
Db 4007 -----YSLRVD-----LRAGDEAVF 4021
QY 1241 HNEKVSVILRGLPNVVTSAISLPNIRKPDGYD--YVAFSKDQYINIDVPSRTARA---IT 1296
Db 4022 AQ-----YDSFQVDSADEYRLHLEGYHGTAGDSMS 4052
QY 1297 TRSGQTLS-----KVWY-NC 1310
Db 4053 YHSGSVFSARDRPNLLISCAVSRYRGAWWYRNC 4086

RESULT 13
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: ventral prostate
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:100347, NCBIP:100348)
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Win
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494
R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PRE>
A:Note: this peptide, designated proline-rich polypeptide V, can be found at several lo
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
```

A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loci
 C;Genetics:
 A;Introns: #status absent
 A;Note: single copy gene with no introns
 C;Superfamily: rat prostatic proline-rich peptides 637K precursor
 C;Keywords: prostate; tandem repeat
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-5761/Product: prostatic proline-rich peptides 637K
 Query Match 12.5%; Score 885.5; DB 2; Length 5762;
 Best Local Similarity 10.6%; Pred. No. 6.9e-10;
 Matches 533; Conservative 172; Mismatches 447; Indels 3865; Gaps 151;

QY	4	KTLPIYLILLVSVFVIQVSSQDLSCAGRCGEGYRDATCNCYNCQ	-----OPTQNEYGQLPNVTVRPVDVA	486
Db	456	ETLPIHEI	-----	51
QY	52	-----HVMCCPDPRVCTAELSCKGRCFESFERGRE	-----	83
Db	487	LTVTSEPVKETESFLAPQEPVHALEYSD	-----VEFFVNEEPPVQAP	531
QY	84	-----CCDAOCKYDKCCPDYESFCAEVKDNKNRTKKP	-----	119
Db	532	ETPGESOFESQLEVPAGATEYD	-----EBF-----KTSATEQEQLAQFPENDEVTVL	578
QY	120	-----TPKP-----PVDDEAGSLDNGDF	-----	138
Db	579	PSNHQAQHSILSNVTQDPLDLITITEKPMEMGTSPVYDAAAAPVEVFLSDQQGVLS	638	
QY	139	-----KVTPDNTSTQH	-----	150
Db	639	QSLPILYDLSQOQYTTGISQSEGEPPTQETPEHSGVMHTEVAQPPGHVETGP	698	
QY	151	-----NKUSTSPKITTAKP	-----INRPSLPNPSD	176
Db	699	PLGHGVHPALQNIILTOYSTPPEKQHSVGVGVFCHLEEFVSVPSPQOENSAMHSVT	758	
QY	177	-----TSKETSLTVNKETTVE	-----	192
Db	759	DMFLSPVDLQTFRSTQSKSYKTIKHEDLALTIPEPSLEDGSIILFPQEDLLQPIDSTG	818	
QY	193	-----TK-----ETTTNKKQTS	-----TDGK	208
Db	819	QGEFHSIKTSLTKPPYVNTKSSAFQETMSETYNSKQVLDSSHLTKTSELPPDYTMGL	878	
QY	209	EKTISAKETQS	-----LEKTSAKDLAPTSKVLAKPT	239
Db	879	EPSLYQOITQSSPKSLHENPKSPFVLKPSAAQALKPRKEKLSPTNNMIPHHSKPLKNM	938	
QY	240	-----	-----	239
Db	939	VTHIPAHKMTVPRQIQEQOGEYTISSNGSFQPLDLEVLITSLGIIPEVKHILPKRTVNPQT	998	
QY	240	-----PRAETTTKGP-----ALTTPE	-----	256
Db	999	YSQVKISHSQHVETQHPNSSETTVQPLDLFAINLQTPKENFAQTQDFTQTMIGPPKE	1058	
QY	257	-----	-----PTPTTPKEPAST	268
Db	1059	VIAQAPHEHGETIPIQDQAEYSTLPTVSVFQPLDQELTITSEAIRPEFPHTVPOQTIIIV	1118	
QY	269	TPKE	-----TIKSAPTTPE	287
Db	1119	HPPEHPLVHSBQVYVQHNPTEAIQPLDLELTITPQETAGELPQTLQDSTTQIIEPP	1178	
QY	288	-----PAPT-----	-----	291
Db	1179	TVVGVPIYIEVTVQTTSDQAEYPPSPVTSFQSLDLELTITSEATRESYHPSLLQQT	1238	
QY	292	-----	-----	291
Db	1239	IVNPEHPLVHSEQVHTQHNLTEATVQPLDPLDITITPQPTTEGELPQTLQDSTQSI	1298	

QY	292	-----TTKSA-----	-----PTTPKE	302
Db	1299	EPPIKWVALVYQEVSDQAEYTTSTSVSFQPLDQELTITSEAIRPEHHTVFPQTTIV	1358	
QY	303	-----PAPT-----TTKEPAPTTPKE	-----	318
Db	1359	HPTKHLPLVHSEQVHTQHNPTEVTVQPLDLELTITPQPTAEGELPQTLQDSTTQIIEPPTV	1418	
QY	319	-----PAPT-----	-----	323
Db	1419	VVGVPYIEVTVQTTSDQAEYPPSPVTSFQSLDLELTITPTEHETITQKTTVFPDPMY	1478	
QY	324	-----TKEPAP-----TTKSAPTTPEK	-----A	343
Db	1479	TDVTLPOQVSVQHLKPTGEGIVQPLDLELTITPQPTBEGELSQTVQEBSTTKQKPEHKEVVA	1538	
QY	344	PT-----TPKK-----	-----	349
Db	1539	PVPYQAVTVPTPSQVQAEYQKSLQPLDLELTITVSEPTKEAYHSTISKNLSAINPQYVHI	1598	
QY	350	-----PAPT-----	-----PKE-----	357
Db	1599	QHENPAEATVQPLDLELTITSSLOPTAEGELLYSMQETVQISEPPKQVTVTPVEYQEA	1658	
QY	358	-PAP-----	-----TTPKEPTTPPE	373
Db	1659	VPAPVQDQAKYPLSSIVLSNLSLQDELTSSELLGEAHLTPDPTWVLPKDKQGIYPDH	1718	
QY	374	-----	-----	373
Db	1719	DHKKHLNLTETVNTQPPHLEHTVQHPTIEERSQSIQKTTQITEPGKVVPLAQESEV	1778	
QY	374	-PAPTTKBPAPTTP	-----	386
Db	1779	TIPMILKETAPTTHSMALQSLDEKLTIHSHSPGWTQHANKESKHTTGKILLDYAE	1838	
QY	387	-----	-----KE	388
Db	1839	PNMEIELKHGLLKTTEATTESENTNQMTSKLQVTLFTQNKKSMLPALVESQESQ	1898	
QY	389	PAPTAPKPK-----APTPK-----BPAPT	-----	408
Db	1899	PPNNMSLQPLDQELTSSQPHGVPHIPNTPEKIYLHYAEP-FTGPFVPEPPELFLKTK	1957	
QY	409	-----TPKE-----PAP-----	-----TTTKE	420
Db	1958	SRPVQGTATQMAASPKEVSRAPENKEAVLSQGEDQDESPPNNMSLOSLOQELTSSQ	2017	
QY	421	PSPTTPKEP-----APT-----TTKSAP	-----	438
Db	2018	PHGWIPHPNTHGKIYLHYAEPPTGPFVEPDLFFLTKTKSKPEVWTLTRTDKSRKEMVS	2077	
QY	439	-----TTTKEP-----	-----	444
Db	2078	QSPKYEEAVLPVHGQGEESRPPNNMSLOSLEQELTSSQPHGVPHPPNTHGKIYLHYA	2137	
QY	445	-APT-----TTKSAP	-----	453
Db	2138	EPPTGPFVEPDLFFLTKTKSKPVGTATRMVKSPEEMVSLDPENKEAVFPAQGEKGES	2197	
QY	454	-----	-----TTP	456
Db	2198	PSSNNMSLOSLEHFMSSQPHGWIHPPKPTDKIYLHYAEPPTGPFVEPDLFFLTK	2257	
QY	457	KEPSPTTTKEPAPTTPKEPAPTTP-----KGPAP-----	-----T	486
Db	2258	SKPVQCHTTEMA-KSPKEMVSQTEYKEAVLSGCEQDESPPNTSLKSLDQEVAMS	2316	
QY	487	TPKEPAPTTPK-----EPAP-----TTTK-KPAPTAP	-----KEP	515
Db	2317	QPHSGVPHPPKTPGKIYLHSTEPFPPVFKPTDILLVTKTKSPAETWTPRDKLKEM	2376	
QY	516	APTPK-ETA-----	-----PTPKKL	531

QY 746 PKKAPPTPE--TPPPTTSE-VSTPTTTKEPTTIHKSPDBESTPSELSAEPFPPKALENSPKX 802
 Db 802 TTTTEKTTKTTTEKTTSESATTTTSEPST-----TEST----- 838
 QY 803 PGVPTTKTAATKPEMTTAKDKTTRDLRTTTPETTTA----- 840
 Db 839 ----TVDTSATTEESSTAETTTTSAB--TSETTTSESAAITGESPENTALQSSSQK 891
 QY 841 -----APMTKETATTTTEKTTTESKITATTTQ-VTST---TTQDTTPFK 879
 Db 892 SEENESSAEPKPGARRDFVFKKHTTVKPAETTSAGVAASITTTTPTTEKTTSTTLETTPIE 951
 QY 880 ITTLKTTTLAPKVT---TTKTTTITTEIMNKPETAKPKDRATNSKATPKPQ----- 929
 Db 952 ATTLNEVTGFAFVGAPVDETTINTLELJSK-----INNQTISQPKPTDISKTD 1001
 QY 930 -----KPTKAPKKTSTKKPTMPRVKPKTPTTPKMTSTMPELN----- 970
 Db 1002 LSSLISGLIGSFTKAPWAPI-----HITDAAFTVATEASLNDGSKKIID 1048
 QY 971 ----PTSRIAEAMLOTTTRENQTSKLVNPKSEBAGAGETPHMLLRPHVFMPEVTP 1027
 Db 1049 EAQPTDEIIRA-----OPTN-----EMDK 1067
 QY 1028 DMDDLPRVFNQGIIL-----NPMI----- 1046
 Db 1068 EMEPEKRIEORIQMQAKRLREELLEKQLOQEIEKARNEMIERKQKMLQOELKE 1127
 QY 1047 -----SDETWICNGKPDVGLTTLRNGTLVAFRGHYFWMLSP 1082
 Db 1128 ABERQORVLEQERLOEQERQLAEKEAELAFG-----SISTTTEASKFKYR----- 1174
 QY 1083 FSPSPAR-----RITFVWGIPIPDITVFRNCBE-GKTFPFKDSQYWRFTNDIKAGYP 1136
 Db 1175 ----LRPAQCAAINKFRVNTDPSMWI--QKNCEFAKRYF-----PEASCP 1216
 QY 1137 KPIFKFGGLTGQIVAAALSTAKYKNWPESVYFF 1169
 Db 1217 -----QIQALI-----ESCFAP 1228

RESULT 15
 A35175
 N:Title: mucin 1 precursor, repetitive splice form A [validated] - human
 N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin
 N:Contains: mucin; polymorphic epithelial mucin (PEM)
 C:Species: Homo sapiens (man)
 C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000
 C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51
 R:Lightenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
 J. Biol. Chem. 265, 5573-5578, 1990
 A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene and
 A:Reference number: A35175; MUID:9020794; PMID:2318925
 A:Accession: A35175
 A:Molecule type: mRNA
 A:Residues: 1-952,1033-1344 <LIG1>
 A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:
 A:Experimental source: Splice form A
 A:Note: GenBank entries HUMEPIA1 and HUMEPIA2 present only the amino-and carboxyl-en
 A:Accession: B35175
 A:Molecule type: mRNA
 A:Residues: 1-19,29-952,1033-1344 <LIG2>
 A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
 A:Experimental source: splice form B
 A:Note: GenBank entries HUMEPIB1 and HUMEPIB2 present only the amino-and carboxyl-en
 R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
 A:Reference number: A35886; MUID:90368715; PMID:1697589
 A:Accession: A35886
 A:Status: not compared with conceptual translation

A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA9876.1; PID:g188870
 A:Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
 R:Ian, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716; PMID:2394722
 A:Accession: A35887
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
 A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
 A:Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated sequence
 R:Wreschner, D.H.; Harsuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.;
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413; PMID:2351132
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRB>
 A:Cross-references: EMBL:X52229; NID:g37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2>
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
 R:Abe, M.; Siddiqui, J.; Kufe, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473; PMID:2597151
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
 R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglut
 A:Reference number: JX0235; MUID:93123189; PMID:1478919
 A:Accession: PX0066
 A:Molecule type: mRNA
 A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037;1038-1057 <MAS>
 A:Experimental source: Gastric carcinoma cell
 R:Zrian-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 336, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
 A:Reference number: S51026; MUID:95080414; PMID:7988707
 A:Contents: annotation
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 Partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
 C:Genetics:
 A:Gene: GDB:MUC1; PUM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q21-1q23
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 A:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
 F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
 F:1-23/Domain: mucin 1 amino-terminal non-repetitive
 F:1-62/Region: signal sequence #link PREA #status predicted <SIGA>
 F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
 F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
 F:1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
 F:138-1017/Region: 20-residue repeats (GSTAPPAHGVTSPDTPAP)
 F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1245-1272/Domain: transmembrane #status predicted <TRM>
 F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.3%; Score 867.5; DB 1; Length 1344;
 Best Local Similarity 21.8%; Pred. No. 7.1e-11;
 Matches 371; Conservative 93; Mismatches 423; Indels 815; Gaps 84;

```

QY 7 PIVLLLLSVFVIQVSSDLSSCAGCGEGYSDATCNCYHQHYMECCPDFREVCYA 66
Db 8 PFLLLLLVLL----- 18
QY 67 ELSCKGRCFESPERGECDDAQCKYDKCCPDYESFCAEVKDNKKNRTKKRPTKPPVV 126
Db 19 ----- 18
QY 127 DEAGSLDNGDKVITPDTSTTQHNKVVSTSPKITTAKPINPRSLPPNSDTSKETSILTVN 186
Db 19 -----TATTAP 24
QY 187 KETTVETKETTNTKOTSDGKEKTTSAKETOSIEKTSAKD-LAPTSKVLAKPTP-KAET 244
Db 25 KRAIV-----VTGSHASSTPGGEKETSATORSVSPSTEKNAVMTSSVLSSHSPGSGS 80
QY 245 TTKGPALTTKEPTPTTPKEPA-----STTP-----KE 272
Db 81 TTQGDVTL-----APAT--EPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134
QY 273 PTFPT-----LKSAPTTKPEPAPTT-----KSAPTTKPEPAPTTKPE-----A 312
Db 135 PAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAP 190
QY 313 PTTKPEPAPTTKPEPAPTTKSAPTTKPEPAPTT-----KKPAPTTKPE----- 358
Db 191 PGT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 247
QY 359 --APTTKPEPTPTT-----KEPAPTTKPE-----APTTKPEPAPTAP----- 394
Db 248 TSAPOTRPAAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 307
QY 395 -----KKPAPTTKPE-----APTTKPEPAPTTKPE-----SPTTKPEPAPTTK----- 435
Db 308 TSAPOTRPAAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPA 363
QY 436 ----SAPTTKPEPAPTT-----KSAPTTKPEPAPTTKPE-----APTTKPEPAP 477
Db 364 AHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 419
QY 478 TTP-----KKPAP--TTP-----KEPAPTTKPE-----APTTKPEPAP 509
Db 420 TAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 479
QY 510 TAPKEPAPTTKETAPTTPKLTPTTPEKLAPTTPEKPAPTTPE-----ELAPTTPEPTP 565
Db 480 TAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 539
QY 566 TTP-----EEPAP--TTPKA-----AAPTTPKEPAPTT-----KEPAPTT 599
Db 540 TAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 599
QY 600 PKEP-----APT-----PKETAP-----TTPKG-----TAPTTLKEPAP 629
Db 600 TAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 659
QY 630 TTP-----KKPAPKELAPTTKPE-----TSITSDKAP--TTPKG-----TAPTTP 669
Db 660 TAPPAHGVTSAPDT--RPAGSTAP-----PAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT 714
QY 670 KEPAPTT-----KEPAP--TTPKG-----TAPTTLKEPAPTT-----KK 703
Db 715 PAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT 774
QY 704 PAPKELAPTTKGTSTSDKAP--TTPK-----ETAPTTKPEPAPTT-----K 747
Db 775 PAPGSTAP--PAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT 833

```

```

QY 748 KPAPTTPEPTTSEVSTPTTTTKEPTT-----IHKSPDESTPESLSEPTPKALENSP 800
Db 834 REAPGS--TAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAP 891
QY 801 KEPGVPTTKTPAA-----TKPEMTTAKOK--TTERDLRTTTPETT-----TAAP 842
Db 892 DTRPAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAP 951
QY 843 KMTKETATTTETTESKITATTQTSTTTQDTTTPKITTLLKTTTLAPKVTTTKTITTT 902
Db 952 -----DTRP----- 955
QY 903 EIMNKPEETAQKDRATNSKATTPEQKTKAPKPTSTTKPKTMRVVKPKTPTTPREM 962
Db 956 -----APGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--APGSTAPPAHG 1006
QY 963 TSTMPELNP-----TSRIAEAMLQTTTRPNQTP--NSKLVEVNPKSDAGAGETPHMLL 1016
Db 1007 VTSAPDT--RPAGSTAPPAHGV--TSAPDNRRPALGSTAPVHNVTISASGSAG----- 1056
QY 1017 RPHVFMPEVTPMDXLPVFNQGIINMPLSDETNICNGKPVGDLTTLRNGTLVAFRGHY 1076
Db 1057 -----SASTLVHNG-----TSARATTTASKS-- 1078
QY 1077 FWMLSPEPSPARRLITEVWGIPSPIDTVTRCCEGKT----- 1115
Db 1079 -----TPFSIPS-----HHSDDTTTLASHSTKTDASTHHSTVPEPLTSSNHSTS 1122
QY 1116 -----PFF-----KDSQVWRFTNDIKDAGYKPIFK--GFGLT- 1147
Db 1123 POLSTGVSTFFLSPHLSNLQFNSSLEDPSTDIYQELQDISMFL--QIYKQGGFLGLSN 1180
QY 1148 -----GOIYVAALSTAKYKNWPESVFFKRG--STQOYIYKOEVPVK----- 1187
Db 1181 IKFRPGSVVVQULTA-----FREGTINVHDVETQFNQYKTEAASRYNLITISDVS 1229
QY 1188 -----CPGER----- 1192
Db 1230 VSDVFFPFSQAQSGAGVPGWGIALLVLCVLVALIYVLIALAVCQCRKNYGLDIFPAR 1289
QY 1193 -----PALNYPVYCEMTQVRRRRRFRERAIQPSQTHIRIQYSP-----ARLAYQDKG 1238
Db 1290 DTYHPMSEYPTY-----HTHGRVVPSSST-----DRSPYEKVSAGNGSSLSYTNPA 1336
QY 1239 VLNHEVKVGSILWRGLPNVVTSA 1260
Db 1337 V-----AATSA 1342

```

Search completed: October 13, 2004, 12:03:35
 Job time : 106.704 secs

This Page Blank (uspto)